American Sociological Review

Official Journal of the American Sociological Association

SCOPE AND MISSION: The *American Sociological Review* publishes work of interest to the discipline in general, new theoretical developments, results of research that advance our understanding of the most fundamental social processes, and important methodological innovations. Like other publications of the Association, emphasis is given to exceptional quality. Unlike the more specialized journals of the Association, the *American Sociological Review's* primary focus is on whatever has the most general bearing on the knowledge of society.

Concerning manuscripts, address: Editor, *American Sociological Review*, Department of Sociology, Ohio State University, 300 Bricker Hall, 190 North Oval Mall. Columbus, Ohio 43210-1353. Phone (614) 292-6200.

Concerning advertising, changes of address, and subscriptions, address: Executive Office, American Sociological Association, 1722 N Street, N.W., Washington, D.C. 20036. Annual membership dues of the ASA, $26–$134 ($16 of this amount may cover a subscription to *American Sociological Review*); Associate, $43; Student Associate, $26. Subscription rate for members, $16; non-members. $38 U.S., $43 elsewhere; institutions, $78 U.S., $83 elsewhere; non-member students, $18 U.S., $23 elsewhere; single issues, $5. New subscriptions and renewals will be entered on a calendar year basis only. *Change of Address:* Six weeks advance notice to the Executive Office, and old address as well as new, are necessary for change of subscriber's address. Claims for undelivered copies must be made within the month following the regular month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

(Articles in the *American Sociological Review* are indexed in the Social Sciences Index, Psychological Abstracts, Sociological Abstracts, Ayer's Guide, University Microfilms, Abstracts for Social Workers, International Political Science Abstracts, United States Political Science Documents, and Current Index to Journals in Education.)

P4620

AMERICAN SOCIOLOGICAL REVIEW

3ø/. ø5

ß-n 3 5

February 1988   Vol. 53, No. 1

Notice to Contributors

STATEMENT OF ASA POLICY ON MULTIPLE SUBMISSION

Submission of manuscripts to a professional journal clearly implies commitment to publish in that journal. The competition for journal space requires a great deal of time and effort on the part of editorial readers whose main compensation for this service is the opportunity to read papers prior to publication and the gratification associated with discharge of professional obligations. For these reasons the ASA regards submission of a manuscript to a professional journal while that paper is under review by another journal as unacceptable.

Section II.B4, ASA Code of Ethics

AUTHOR'S RESPONSIBILITY: *Manuscripts which do not adhere exactly to these suggestions will be immediately returned for revision. Only articles in the form specified below will be sent out for refereeing.*

The third edition of the *Publication Manual* of the American Psychological Association (available from A.P.A., P.O. Box 2710, Hyattsville, MD 20784) is very helpful in clarifying headings, abbreviations, table preparation, and the like.

Since manuscripts are evaluated as anonymous submissions, please keep identifying material out of the manuscript. Attach a cover page giving authorship, institutional affiliation, acknowledgments, and the date the article is submitted. Provide only the title as identification on the manuscript and abstract. Every effort should be made by the author to see that the manuscript itself contains no clues as to the author's identity.

Submission and Preparation of Manuscripts

1. Submit four (4) copies and retain the original for your files. Copies may be Xerox, mimeograph, or multilith, but not carbons. Manuscripts must be accompanied by a $10.00 manuscript submission fee, payable to the American Sociological Association.

2. Enclose a stamped self-addressed postcard for acknowledgment of manuscript receipt. Since manuscripts will not be returned by reviewers to the editorial office, the editorial office will not return them to authors.

3. All copy must be *typed, doublespaced* (including indented material, footnotes, and references) on 8½ by 11 inch white opaque paper. Line must not exceed six (6) inches. Margins must be a minimum of 1¼ inch on left and foot.

4. Include four (4) copies of an abstract of *no more* than 150 words.

5. Tables should be numbered consecutively throughout the article and typed on separate sheets at the end of the manuscript. Insert a location note at the appropriate place in the text, e.g., "Table 2 about here." Each table must include a descriptive title and headings to columns. Gather general footnotes to tables as "*Note:*" or "*Notes:*," and use *a, b, c,* etc., for specific footnotes. Asterisks * and/or ** indicate significance at the 5 percent and 1 percent levels, respectively.

6. Illustrations submitted with the final draft must be of professional quality, ready for reproduction, executed on white paper or vellum, in black ink, with clear, medium weight, black lines and figures. All lettering on figures should be executed by an artist in pen and ink, by means of dry transfer letters, or by applying typeset material to the prepared artwork. Typewritten lettering should not appear in illustrations. Figures should be capable of legible reduction to a width size no larger than 5³⁄₁₆ inch (full page) or 2⁹⁄₁₆ inch (one column); they should be numbered consecutively, and the number and author's name should be penciled lightly on the back of each. All illustrations must have captions, which should not appear on the artwork but should be typed, double-spaced, on a sheet at the end of the manuscript.

7. Equations must be typed and important displayed equations identified by consecutive arabic numbers in parentheses on the right. Expressions should be aligned and compound subscripts and superscripts clearly marked if there is any potential for confusion. Indicate boldface characters by drawing a wavy line (a) under them; a single underline (a) means italic to a printer. Clarify all symbols with notes in the margins of the manuscript. Circle these and all other explanatory notes not intended for printing.

8. Three kinds of footnotes are possible each serving a different purpose:

   a. Content footnotes. Content footnotes are explanations or amplifications of the text. Because they are distracting to readers and *expensive* to include in printed material, an author should include important information in the text and omit irrelevant information. *Content footnotes will not be allowed generally.*

   Rather than footnoting long or complicated material, such as proofs or derivations unnecessary to the text, consider (i) indicating in a *short* footnote that the material is available from the author, (ii) depositing the material in a national retrieval center and including an appropriate footnote, or (iii) adding an appendix. If an appendix is used, the reference in text should read, for example: "(see Appendix A for complete derivation)."

   Text footnotes should be numbered consecutively throughout the article with superscript arabic

numerals. If after a footnote occurs it is later mentioned, use a parenthetical note "(see note 3)," rather than the superscript number.

b. Reference footnotes. Footnotes are used for reference purposes only to cite material of limited availability. Acceptable reference footnotes include (i) legal citations, which should follow the footnote style of *A Uniform System of Citation* (1967) published by the Harvard Law Review Association, (ii) copyright permission footnotes, (iii) unpublished works and works in progress.

c. Table footnotes. Table footnotes are appended *only* to a specific table. Footnotes to a table should be *lettered* consecutively *within* each table with superscript lowercase letters. (See 5.)

9. Acknowledgments, credits, and grant numbers are placed on the title page with an asterisk.

Reference Format

A. *In the text:* All source references are to be identified at the appropriate point in the text by the last name of the author, year of publication, and pagination where needed. Identify subsequent citations of the same source in the same way as the first. Examples follow:

1. If author's name is in the text, follow it with year in parentheses [". . . Duncan (1959) . . ."].

2. If author's name is not in the text, insert in parentheses the last name and year [". . . (Gouldner 1963) . . ."].

3. Pagination follows year of publication after a comma [". . . Kuhn (1970, p. 71)."].

4. Give both last names for dual authors. Give all last names on first citation in text for more than two authors; thereafter use "et al." in the text. When two authors have the same last names, include initials in the text. For institutional authorship, supply minimum identification from the beginning of the complete citation [". . . (U.S. Bureau of the Census 1963, p. 117) . . ."].

5. Separate a series of references with semicolons and enclose them within a single pair of parentheses [". . . (Burgess 1968; Marwell et al. 1971, pp. 386–87; Cohen 1962) . . ."].

B. *In the appendix:* List all items alphabetically by author and, within author, by year of publication in an appendix titled "REFERENCES." The reference appendix must be complete and include all references in the text. The use of "et al." is not acceptable in the appendix; list the names of all authors using full first names. (See A.4. for text format.)

If there is more than one reference to the same author and year, distinguish them by the letters a, b, etc. added to the year [". . . (Levy 1965a, p. 331) . . ."].

The first letter of each word in an article title should be capitalized. Titles of books and journals are printed in italics, so each word of the title should be underlined.

Give the publisher's name in as brief a form as is fully intelligible. For example, John A. Wiley and Sons should be "Wiley."

If the cited material is unpublished, use "forthcoming" with name of journal or publisher; otherwise use "unpublished."

Examples follow:

1. *Books:*   Mason, Karen O. 1974. *Women's Labor Force Participation and Fertility.* Research Triangle Park, NC: National Institutes of Health.

         U.S. Bureau of the Census. 1960. *Characteristics of Population.* Vol. 1. Washington, D.C.: U.S. Government Printing Office.

2. *Periodicals:*   Conger, Rand D. Forthcoming. "The Effects of Positive Feedback on Direction and Amount of Verbalization in a Social Setting." *Pacific Sociological Review.*

         Goodman, Leo A. 1974a. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.

         _____. 1974b. "The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–1259.

3. *Collections:*   Clausen, John A. 1972. "The Life Course of Individuals." Pp. 457–514 in *Aging and Society,* vol. 3, *A Sociology of Age Stratification,* edited by M.W. Riley, M. Johnson, and A. Foner. New York: Russell Sage.

         Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165–90 in *Annual Review of Sociology,* vol. 1, edited by A. Inkeles, J. Coleman, and N. Smelser. Palo Alto, CA: Annual Reviews.

See 1986 and later issues for further examples.

CONTRIBUTORS TO THIS ISSUE

■ PAMELA E. OLIVER (The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II.) is Associate Professor of Sociology at the University of Wisconsin. She has recently published other articles on formal collective action theory and also publishes in the related area of collective behavior and social movements. She is this year's chair of the American Sociological Association's section on Collective Behavior and Social Movements. GERALD MARWELL is Professor of Sociology, University of Wisconsin-Madison. Besides this multi-article project with Pamela E. Oliver on "Theory of the Critical Mass," he is working on a book titled *After Activism: Sixties Civil Rights Workers Turn Forty.* The first article from this panel research project has recently appeared in *Public Opinion Quarterly.*

■ CHARLES N. HALABY (Action and Information in the Job Mobility Process: The Search Decision) is Professor of Sociology, University of Wisconsin-Madison. His research interests include labor markets, mobility, industrial sociology, complex organizations, and statistical methodology. Most of his recent research is about the historical evolution and current state of the employment relation.

■ THOMAS A. DiPRETE (Gender and Promotion is Segmented Job Ladder Systems) is Assistant Professor, Department of Sociology, University of Chicago. He has recently completed a book-length manuscript on the development and impact of a tiered personnel system in the federal civil service. He is currently engaged in a collaborative project with David B. Grusky to specify and explain recent trends in status and income attainment for American men and women. Recent papers include "Horizontal and Vertical Mobility in Organizations" (forthcoming in *Administrative Science Quarterly*) and "The Upgrading and Downgrading of Occupations" (forthcoming in *Social Forces*). WHITMAN T. SOULE is a consultant involved primarily with civil rights litigation. He is currently a doctoral candidate at Northwestern University researching economic sectoral influences on the division of labor and stratification among scientifically and technically trained workers in the U.S. labor force from 1960 to 1984.

■ ROGER FINKE (Religious Economies and Sacred Canopies: Religious Mobilization in American cities, 1906), Assistant Professor of Sociology, Loyola University of Chicago, is at work on a book analyzing 200 years of development and change in the American religious economy, giving special emphasis to interdenominational competition. RODNEY STARK, Professor of Sociology and Comparative Religion, University of Washington, has just published *A Theory of Religion* (with William Sims Bainbridge) and a series of papers on the ecology of deviance.

■ EDWARD N. MULLER (Democracy, Economic Development, and Income Inequality) is Professor of Political Science at the University of Arizona. He is currently completing a book manuscript, *Dependence and Democracy,* which evaluates a variety of explanations for the stability and breakdown of democracies in the world since 1950. Other current research includes an international panel study to be conducted in Israel, Peru, and West Germany during 1987–90, testing rational choice models of collective political action.

■ WILLIAM BRUSTEIN (The Political Geography of Belgian Fascism: The Case of Rexism) is Associate Professor of Sociology, University of Utah. His areas of interest are theory, political and historical sociology, and social movements. His book, *The Social Origins of Political Regionalism: France, 1849–1981* (Berkeley: University of California Press, 1988), is forthcoming. He is currently working on a study of the political geography of fascist party membership in Italy and Germany during the interwar period.

■ D. WAYNE OSGOOD (The Generality of Deviance in Late Adolescence and Early Adulthood) is Assistant Professor of Sociology, University of Nebraska, Lincoln. His current research interests include interpersonal influence and deviance, the cessation of illegal behavior among serious offenders, and the impact of role transitions on deviance. LLOYD D. JOHNSTON is a Program Director and Research Scientist at the Survey Research Center, Institute for Social Research, The University of Michigan, where he is principal investigator on the ongoing national research project entitled *Monitoring the Future.* He has written numerous articles, chapters, monographs, and books—largely in the field of substance abuse—including *Drugs and American Youth* and *National Trends in Drug Use and Related Factors Among American High School Students and Young Adults, 1975–1986.* He has also been involved in the development of foreign and multi-national studies in the substance abuse area. PATRICK M. O'MALLEY is an Associate Research Scientist in the Survey Research Center, Institute for Social Research, The University of●

iv

Michigan. He is coauthor, with Johnston and Bachman, of *National Trends in Drug Use and Related Factors Among American High School Students and Young Adults, 1975–1986,* published by the National Institute on Drug Abuse, 1987. JERALD G. BACHMAN a Program Director and Research Scientist at the Survey Research Center, Institute for Social Research, University of Michigan, has directed a program of research on youth and social issues for more than 20 years. Publications include *The All-Volunteer Force,* five volumes in the *Youth in Transition* series, and numerous journal articles. Current research interests are focused on drug use and its correlates among youth and young adults.

■ DOUGLAS A. SMITH (The Social Organization of Self-Help) received his Ph.D. in sociology from Indiana University in 1982 and is Associate Professor in the Institute of Criminal Justice and Criminology at the University of Maryland. His recent publications concern police handling of interpersonal violence, ecological models of criminal victimization, and issues in plea bargaining. CRAIG D. UCHIDA received his Ph.D. in Criminal Justice from the State University of New York at Albany in 1982. Currently, he is an Assistant Professor in the Institute of Criminal Justice and Criminology at the University of Maryland. His recent publications deal with the history of police, the use of law enforcement statistics, and the impact of a Supreme Court decision on the police and the courts.

■ NANCY J. DAVIS (Class Identification of Men and Women in the 1970s and 1980s) is Associate Professor of Sociology at DePauw University. Her research has been in the area of women's studies, including the influence of public policy on women's lives, women's education after marriage, and factors affecting childlessness among women in the early 20th century. She is currently studying unemployment and employment in late 19th century America. ROBERT V. ROBINSON is Associate Professor of Sociology and Director of the Institute of Social Research at Indiana University. A past editor of the series *Research in Social Stratification and Mobility,* he recently completed a study of class mobility in several advanced industrial and Third World countries. He is currently working on a social historical study of work and family life in Indianapolis from 1850 to 1930.

■ FRANKLIN D. WILSON (Aspects of Migration in an Advanced Industrial Society) Professor, Afro-American Studies and Sociology, University of Wisconsin-Madison, is continuing his work on migration and metropolitanization. His current work involves testing a life cycle model of metropolitan development using data for the U.S. and three developing countries.

■ NEIL G. BENNETT (Commitment and the Modern Union: Assessing the Link Between Premarital Cohabitation and Subsequent Marital Stability) is an Assistant Professor in the Department of Sociology at Yale University and a Faculty Research Fellow at the National Bureau of Economic Research. Much of his current work addresses issues related to the sociology of marital formation and dissolution. Other work focuses on the development of formal demographic models, their application to mortality estimation in the U.S. and less developed nations, and policy issues surrounding health care in the U.S. ANN KLIMAS BLANC is a demographer with the Demographic and Health Surveys Program at the Institute for Research Development/Westinghouse. Her current research focuses on the relationship between women's status and reproductive behavior in developing countries. DAVID E. BLOOM is a Professor in the Department of Economics at Columbia University and a Research Associate at the National Bureau of Economic Research. His recent research deals with the demography and economics of the family, the effects of population growth on labor markets in developed and developing countries, the arbitration of disputes, immigrant labor market assimilation, and the economic consequences of AIDS.

■ LOWELL L. HARGENS (Scholarly Consensus and Journal Rejection Rates; Further Evidence on Field Differences in Consensus from the NSF Peer Review Studies) is Professor of Sociology at the University of Illinois at Urbana-Campaign. He is extending the research reported here by analyzing data for three scientific fields as part of a comparative study of the manuscript review process. With Andrew Pickering and Robert Alun Jones, he coedits the annual series *Knowledge and Society.*

■ STEPHEN COLE (Do Journal Rejection Rates Index Consensus?) is Professor of Sociology at the State University of New York at Stony Brook. He is currently completing two monographs: *Social Influences on the Cognitive Content of Science* (with J.R. Cole) and *Culture and the Occupational Attainment of Women* (with R. Fiorentine). GARY SIMON is an Associate Professor in the Department of Statistics and Operations Research at New York University. He is currently doing research on missing data and censored observations. NATHAN R. COLE is Vice President for Arts and Sciences and Professor of Sociology at Columbia University. He is completing two books: *Health Risks in the Media* and *Women in the Pursuit of Science* (with H. Zuckerman and J. Bruer).

GUEST EDITORIAL

DATA SHARING AS A PUBLIC GOOD*

Below is the second in a series of guest editorials that discuss important ASA publications issues. The editorial does not represent official ASR policy, which is made by the Publications Committee. Ed.

Robert M. Hauser's editorial in *ASR* (December 1987) focused on two needs: standardized data reporting in research publications; and institutional data sharing in the social sciences. I don't disagree with anything in Hauser's editorial. (In fact, Hauser has several times shared data with me.) I do wish to explore further the issue of institutionalized data sharing because it raises some potential problems of logistics and incentives (alluded to by Hauser).

Data sharing promotes cumulative progress within a discipline by encouraging researchers to use common research tools and to expand on previous studies. However, many areas of sociology do not yet have widely shared procedures and standards for data collection, index construction, and the like. Accordingly, "documentation" is more complex and idiosyncratic when there are fewer institutionalized criteria for constructing data bases and variables. Moreover, in the less developed areas of sociological inquiry, other users are likely to need much more assistance from the creator of a data base than in areas (e.g., labor-market attainment) in which samples, variables, and indices are more institutionalized. I know Hauser believes that data sharing is needed *precisely* to help advance our discipline. However, without adequate resources and incentives to support documentation, dissemination, and assistance to other users, I worry that mandated data sharing could place a very heavy burden on many researchers—particularly those collecting novel kinds of data or working in uncharted areas.

This raises a thorny logistical issue: exactly what data are to be shared? Documentation and standardization of data reporting make obvious sense and must be encouraged; it is certainly valuable to have access to other researchers' computer tapes or moments matrices for replication and model respecification. However, much of sociology is "pre-paradigmatic," and often the most interesting and crucial questions concern how an investigator specified a model or measured a given construct. Especially when

sociologists are analyzing historical or archival materials, field or interview notes, data generated through content analyses, and the like—which I think they are doing increasingly—the "raw" data may be more valuable than the "cooked" to other investigators. A published correlation matrix for the variables analyzed by an author does not permit me to test competing hypotheses involving other variables in the data set that the author did *not* use. Nor does a floppy diskette containing someone's codings of organizational structure for a sample of firms allow me to examine whether, given the same information, I would have coded things similarly. Are authors expected to report (or share) *all* available information, including variables or cases that were analyzed but not reported in an article (e.g., because the results weren't "significant"), or only the information needed by others to replicate the published results? Moreover, in cases where researchers coded variables themselves, should we expect sharing of the source documents (e.g., organization charts or field notes)? Precisely where does one draw the line between so-called qualitative and quantitative data in data sharing? Even if we surmount the logistical problems that Hauser mentions, such as ensuring the accuracy of shared files and compatibility of data files and computer systems (which will be difficult), other problems are likely to vex us.

Permit me a personal illustration, which allows me to expiate some guilt about not sharing data as much as I should have. William Bielby and I invested roughly three years in coding information on thousands of jobs and hundreds of organizations analyzed by the U.S. Employment Service. This project involved photocopying roughly 200,000 pages of source documents and coding them into machine-readable form. Some documents contained qualitative information about organizational arrangements, which was content analyzed. Other forms consisted of more standardized information, such as education and experience requirements for jobs, but necessitated complex coding rules to effect comparability over time. To this information we merged data obtained from public sources, such as industry censuses and the *Dictionary of Occupational Titles*. We•

then spent another year and a half assembling and cleaning data files and creating measures for statistical analyses.

The result was a complex and unorthodox amalgam of multilevel data sets and raw documents. Staggering resources would be required to document all these files completely so that they could be used independently by any investigator, to disseminate these files to the parties who have requested them, and to answer users' questions. There was no way to have envisioned the resource requirements before we got our hands dirty. (In fact, like others, our NSF grant was reduced midstream.) As a result, while Bielby and I have done our best to document our efforts and to satisfy specific data requests, we never produced codebooks and data extracts that could be provided to users "off the shelf." I suspect our experience is, regrettably, all too typical.

This highlights some incentive problems associated with mandated data sharing. (I am leaving aside the obvious problem of researchers' desires to preserve monopoly power.) Such a policy represents a public good, involving the well-known problems about how to fund such goods. For instance, an individual with an NSF grant to merge Census industry characteristics to the Panel Study of Income Dynamics (PSID) can "free-ride" on the documentation efforts of the institutions that have already created these primary data sources. In contrast, an individual who collects original data pays a double tax: having to field the data collection and having to do the original documentation as well. (The taxes are even steeper against individuals who collect novel sorts of data, since there are fewer protocols for documentation and users will require more guidance from the creator of the data base. The taxes are also steeper against those at resource-poor institutions.)

Without subsidies, a data-sharing requirement encourages researchers to use data sets for which the requisite costs have already been borne by institutions, such as NORC's General Social Survey, the PSID, and Census public use tapes. While it is certainly good to encourage researchers to work on common data sets, it is not obvious to me that the most innovative and worthwhile research in our discipline nowadays employs these standard data sets. It is worth thinking about whether mandated data sharing might affect the distribution of research efforts.

As Hauser notes, resources must be made available if such a policy is implemented. But how? NSF now requires data sharing after a certain period of time. This makes sense to me, but we all know that there is sharing and there is SHARING. When creative researchers must finance privately the clerical, administrative,

and consultative tasks required by a policy of mandatory data sharing, this imposes large opportunity costs, not only on the researchers, but also on the discipline.

The answer, of course, is not to absolve researchers from the obligation to document and disseminate data, but rather to provide the requisite resources and incentives. I would feel more comfortable about the NSF policy if it were accompanied by a statement about the Foundation's interest in considering supplemental "add-on" requests for funding data sharing either during the course of a grant or at its conclusion, after a researcher has obtained a better feel for what the task will require. To surmount the incentive problems I have described, funding agencies will also require keen sensitivity to differences across research projects, career stages, and academic institutions in the demands that documentation, data sharing, and assisting colleagues place on investigators.

Another enforcement mechanism I have recently heard discussed involves having journal editors verify that certain conditions of data accessibility have been satisfied by authors when accepting manuscripts for publication. I do not believe this will be effective, given all of the logistical issues involved. Journal editors and reviewers are hardly in a position to monitor quality, format, timeliness, or completeness of data sharing. Moreover, if specific sociology journals were to take on this role, there could be incentives for investigators either to send manuscripts elsewhere or to delay submissions until ready to disseminate the data, neither of which I welcome.

Instead, I am more sympathetic to the call for standards and mechanisms promulgated by the ASA, where the Committee on Professional Ethics or a separate body might assist individuals who have been stymied in accessing someone's data and might sanction data hoarders. Another possibility is that we begin institutionalizing tabulations of what might, following sports terminology, be called research "assists": publications that resulted from a researcher documenting and sharing data with other investigators.

Information on grants and *Social Science Citation Index* counts is routinely considered in hiring and promotion decisions. Information on one's track record in facilitating research by other investigators should be equally valuable, not only for assessing contributions to the profession, but also for predicting contributions to graduate student training and the intellectual quality of life in a department. We know from organizational research that information, once collected, tends to be ascribed importance. If we really wish to foster the norm of documentation

and data sharing, we must somehow begin tabulating research "assists" as a routine part of the academic scorecard.

I've highlighted some logistical and incentive problems associated with a data-sharing policy that might produce unintended consequences. I very much hope these caveats will not be interpreted as opposition to a policy that clearly promotes more open, accurate, cumulative, and professionally responsible inquiry. Instead, let's see them as challenges to surmount as we pursue the worthwhile objective of better science.

James N. Baron
Stanford University

MANUSCRIPTS FOR THE
ASA ROSE SOCIOLOGY SERIES

Manuscripts (100 to 300 typed pages) are solicited for publication in the *ASA Arnold and Caroline Rose Monograph Series*. The Series welcomes a variety of types of sociological work—qualitative or quantitative empirical studies, and theoretical or methodological treatises. An author should submit three copies of a manuscript for consideration to the Series Editor, Professor Ernest Q. Campbell, Department of Sociology, Vanderbilt University, Nashville, TN 37235.

viii

THE PARADOX OF GROUP SIZE IN COLLECTIVE ACTION: A THEORY OF THE CRITICAL MASS. II.*

PAMELA E. OLIVER   GERALD MARWELL

University of Wisconsin

Many sociologists incorrectly believe that larger groups are less likely to support collective action than smaller ones. The effect of group size, in fact, depends on costs. If the costs of collective goods rise with the number who share in them, larger groups act less frequently than smaller ones. If the costs vary little with group size, larger groups should exhibit more collective action than smaller ones because larger groups have more resources and are more likely to have a critical mass of highly interested and resourceful actors. The positive effects of group size increase with group heterogeneity and nonrandom social ties. Paradoxically, when groups are heterogeneous, fewer contributors may be needed to provide a good to larger groups, making collective action less complex and less expensive.

Empirical researchers have often found that the size of a group is the best predictor of its level of collective action. Spilerman (1970, p. 654) summarized his analysis of the black riots of the 1960s: "[T]he larger the Negro population, the greater the likelihood of the disorder. Nothing else appears to matter." Scott and El-Assal (1969) found that size of student body was the only significant predictor of demonstrations and other disturbances on college campuses. Interpreting their results, Marwell (1970) argued that the simplest theory would assume that

> a given proportion of students [at all schools] are ready to stage a demonstration in response to certain types of events but . . . this proportion is small. Given that a demonstration is a collective event, it takes some minimum number of such students to get a demonstration off the ground. The larger the university, the greater the chance it has to get a minimum number. (p. 916)

Very large constituencies such as Afro-Americans or women have given rise to much larger social movements in the United States than small constituencies like Armenian-Americans or paraplegics. These empirical findings make a great deal of sense, since larger groups have more resources and more people who might contribute them for collective action.

Nevertheless, many believe on theoretical grounds that it is more difficult for larger groups to sustain voluntary collective action. The major source of this belief is Mancur Olson's *Logic of Collective Action* (1965). Hardin describes

Olson's "central conclusion" as "large groups will fail; small groups may succeed (1982, p. 38)."[1]

When theory conflicts with empirical research, the problem usually lies with the theory. Hardin calls Olson's "group size" assertion "the most controversial issue in the contemporary literature on collective action." As he and others have shown, Olson's argument, which seems so plausible at first, does not stand up to close technical analysis. We begin by reviewing the key issue in this dispute, effects of jointness of supply on the size argument.[2] This review suggests a paradox to which we next turn: providing a collective good to a larger interest group may require *fewer* individual contributors. The final section discusses the implications of this paradox by considering it in relation to the social processes underlying the organization of collective action and the conceptualization of the "group" in collective action theory.

Many sociologists have believed that instrumentalist assumptions must be abandoned to account for the obvious inconsistencies between the real world and Olson's "group size" argument, but our analysis of the paradox of group size stays within the instrumentalist framework. We assume that decisions are made

---

[1] In the social psychological literature, "diffusion of responsibility" findings also suggest that the presence of others reduces an individual's propensity to assist someone in need (Piliavin, et al. 1981, pp. 120–32; Latane and Nida 1981).

[2] In an earlier draft, we presented a much more abbreviated review of the literature in this area, and found that reviewers and colleagues who were obviously well-read sociologists nevertheless thought that we were misunderstanding Olson. Thus, we feel it is appropriate to provide a thorough treatment of this issue in a sociological journal, even though the relevant arguments have been published by scholars in other disciplines.

---

consciously with attention to costs and benefits, and that resources are limited. This is a correct, although partial, description of the behavior of people who care about some collective good, have limited resources, and want to "spend wisely" in pursuit of the good. We do not say that people always act out of instrumentalist motives; rather, we show what the consequences will be when they do. We assume that each person can be described by a relatively fixed "potential contribution level." It does not matter to our analysis what individual motives generate these levels, or whether they are the same or different for different people. Solidarity, altruism, or personal morality may all affect the potential contribution level, as may "external" factors such as the intensity of a propaganda campaign or a conversion process.[3]

JOINTNESS OF SUPPLY AND THE GROUP SIZE ARGUMENT

A *public good* is defined by its nonexcludability: if any one group member consumes it, it cannot feasibly be withheld from other group members (Olson 1965, p. 14). Here, a "group" is all individuals in a relevant population who have a positive interest in the good.[4] Anyone who wants to enjoy a public good must be prepared to provide it to everyone in the group.

For Olson, groups come in three theoretically different sizes: *small* or *privileged*, in which some individual may have enough interest in the collective good to provide some level of it himself; *moderate*, in which no individual can provide a significant portion of the good himself, but some individuals can make a "noticeable" difference in the level of provision of the collective good, i.e., affect it enough that it seems to have increased a small amount; and *large*, in which no individual can make even a noticeable difference (p. 44). Although he seems to *define* a large group as one in which no contribution is noticeable (e.g., p. 45), Olson notes that this would be tautological (pp. 48–9n), and recasts his position as "the (surely reasonable) empirical hypothesis that the total costs of the collective goods wanted by large groups are large enough to exceed the value of the small fraction of the total benefit that an individual in a large group would get" (p. 49n). Hence, Olson argues, no rational individual in a large group would ever contribute towards the provision of a public good.

Although this may seem a reasonable empirical hypothesis, it has been well established that Olson's group size argument is either tautological or wrong. Although a variety of critical issues have been raised (Chamberlin 1974; Frohlich and Oppenheimer 1970; McGuire 1974; Oliver 1980), jointness of supply is most important. (See Hardin 1982, pp. 38–49 for a very thorough treatment of this issue.)

A good with *jointness of supply* costs the same no matter how many people "enjoy" it.[5] The classic jointly supplied good is a bridge, which has a fixed cost regardless of how many people use it. The cost of defending a border is roughly the same regardless of how many people are protected within that border. A special interest tax loophole may involve the same lobbying costs whether it benefits one company or a thousand. Jointness of supply can be a matter of degree. A classic private good has zero jointness of supply and a cost which is proportional to the number who enjoy it. A good with pure jointness of supply has all fixed costs and no proportional costs. Between these extremes lie economies of scale, costs that rise less than proportionately with the number who enjoy a good.

Olson discusses jointness of supply in a footnote:

at least one type of collective good considered here exhibits no jointness whatever, and few if any would have the degree of jointness needed to qualify as pure public goods. Nonetheless, most of the collective goods to be studied here do display a large measure of jointness (p. 14n)

Despite this, Olson never discusses how jointness of supply would affect his group size argument.

To appreciate the significance of jointness of supply for the group size argument, it is crucial to recognize that the relevant cost for collective action is that borne by the collective actors. Even though a tax loophole that applies to many will cost taxpayers more than one that applies to few, what matters is how much the *lobbying* will cost. The cost of cleaning up pollution is roughly proportional to the number of polluters, but the cost of obtaining laws requiring polluters to clean up their own messes is not. Government policy as a public good usually has high jointness of supply. An interest group or social movement campaigns for legislation of benefit to them, but their costs are unaffected by the

---

[3] We are not claiming that this is the *best* way to capture such feelings, simple that they *can* be incorporated into an instrumentalist model.

[4] This definition of "group," which is fairly standard in economics, is different from that most common in sociology, and is closer to the idea of an "interest group" or "beneficiary constituency" (McCarthy and Zald 1977).

---

[5] See Samuelson (1954) and Head (1974) for general treatments of public goods.

existence of others who would also be benefited by the legislation.

Zero Jointness of Supply: Group Size Has Negative Effect

Olson's group size argument is clearly correct *only* when the good has zero jointness of supply, i.e., when the cost of providing the good is proportional to the number who share in it. Consider an example loosely drawn from our experience. Imagine that wiring arrangements and university regulations require that all terminals for a departmental computer be placed in public access space and be available to all members of the department. If 50 percent of the department's members want to work at a time, 5 terminals are necessary to provide perfect computer access in a 10-person department, and 50 terminals are necessary for a 100-person department. Suppose individual faculty members are encouraged to buy terminals for the department from their research grants. One terminal is $\frac{1}{10}$ of the number necessary in the smaller department, but only $\frac{1}{100}$ of the number necessary for the larger department. The individual who buys a terminal raises her own ability to work on the computer whenever she wants to by .1 in the smaller group (since all 10 members have equal access to the terminal), but by only .01 in the larger group, where all 100 members have equal access. If a terminal costs $500, an expected-value maximizer would have to value computer access at more than $5000 to be willing to buy a terminal for the small department, which is bad enough, but would have to value access at $50,000 to be willing to buy a terminal for the larger community. Every increase in group size would lower the expected value of a contribution of a given size or, alternately, raise the price of a given level of provision of the good in terms of individual benefit.

If the cost of a nonexcludable good increases proportionately (or more) with the number who enjoy it, larger groups are much less likely to be provided with the good than smaller groups. This is clearly the situation Olson has in mind in his analysis, and there is nothing wrong with his logic. The trouble is that few collective goods meet this condition. Most goods with no jointness of supply are also quite excludable. We had to invent a nonexcludability constraint in our example. Olson's major example is a group of businesses joining together to restrict production in a perfectly competitive market. This is hardly the kind of "collective action" that interests sociologists and political scientists.

Pure Jointness of Supply: Positive Effect of Group Size

The opposite result obtains when goods have pure, or complete, jointness of supply. Then, larger interest groups are much more likely to have a "critical mass" (Oliver, Marwell, and Teixeira 1985) of people willing to provide the collective good. For any individual deciding whether to contribute to a collective good with pure jointness of supply, it is irrelevant how many others might share in the good. Individuals will provide the good if their own benefit from the good outweighs its cost. We may use another example from our own experience. In the course of this research, we had to purchase a simulation compiler. The license for the compiler is for the whole computer, but we could have placed the compiler in a public access file or in a private file available only to ourselves. Obviously, the fact that others can use the compiler is totally irrelevant to the benefits we obtain from it, so we opted for public access. Because this good has pure jointness of supply, we had absolutely no reason to withhold it from others.

In general, the irrelevance of group size for individuals' decisions when there is pure jointness of supply translates into a positive effect of group size on (1) the probability that *someone* in a group will provide the good, and (2) the total amount of contributions from the group. The only exception to this rule is the extreme (but common) case in which the cost of a collective good is so high relative to the interest and resource distributions of those interested in it that no one in the group is willing or able to provide the good. In this case, group size is irrelevant to the outcome. *There really is a dilemma of collective action for public goods, but the dilemma adheres to the high cost of providing them, not to the number who share in them.*

To illustrate this point, let us look at a case at the opposite extreme. When the cost of the collective good is very low relative to the group's interest, the group's size will have little effect on the probability that *somebody* in the interest group will contribute. But group size will have a positive effect on the total number of contributors and the total amount of their contribution. Consider the problem of providing the collective good of calling the power company to report an outage. Making a phone call entails some cost and will benefit others who have not called. There are doubtless many free riders, and even a "diffusion of responsibility" effect of people assuming that someone else is making a call that they would be perfectly willing to make. Nevertheless, someone nearly always calls, whether the affected group is big

or small. In fact, there are almost always quite a
few calls made about any particular outage, and
the number of calls is usually greater the more
people are affected. Although a higher *propor-
tion* of smaller affected populations may call,
there is a greater total number of calls, that is,
more overall collective *action,* in a larger group.

What about intermediate goods, whose cost is
low enough that someone might be willing and
able to pay for them, but high enough that this
willingness and ability is a relatively rare trait?
Continuing with computer examples, consider
the problem of paying the fairly large cost of
linking a computer installation to a world-wide
communication system such as BITNET. Users
at each installation vary in their interest (e.g.,
how much they collaborate with people in other
countries) and in their resources (e.g., discre-
tionary grant or overhead funds). Only a few
users combine high interest in BITNET with
large discretionary funds. For the sake of a
model, we assume that the whole fee must be
paid out of one user's fund, that only a small
proportion of all computer users combine high
interest with large funds, and that users are
randomly distributed across installations.

We can model this as a large population from
which samples of various sizes are drawn.
Imagine, for example, a distribution of interest
(collaboration) and resources (grants) such that
the probability is .01 (i.e., 1 in 100) that an
individual from that distribution would be
willing and able to provide the collective good
(pay the fee). The probability that a sample
(installation) of size n will have at least 1 person
who exceeds a threshold with probability p is 1
$- q^n$, where $q = (1-p)$, while the expected
number who will exceed that threshold is np. If
the installation size is 10, the expected number
of purchasers when p is .01 is .1, and the
probability that someone will purchase BITNET
rounds off to about .1. That is, only 10 percent
of installations of size 10 would be expected to
connect to BITNET. If the installation size is
100, the number who can be expected to
contribute is 1, and the probability that at least
one will do so is about .6, so that 60 percent of
all installations of size 100 would be on
BITNET. But if the installation size is 1,000,
the expected number of users who have both
interest and funds is 10, and it is virtually
certain that there will be at least one person who
will pay for the connection.

*Interactions: Group Size and Economies
of Scale*

Most real cases lie between the extremes of pure
and no jointness of supply. They exhibit partial
jointness of supply, or economies of scale, in
which the cost of the collective good rises less

than proportionately with the number who enjoy
it. In these intermediate cases, the amount of
collective action as a function of group size
depends on the interaction between the cost
function for the collective good and the
distribution of potential contribution levels
among members of the group.

These interactions are always specific to a
particular case, but we can identify the two
important principles that govern them. First, the
more the cost function for the collective good
approximates jointness of supply, the more
likely group size is to have a positive effect on
the provision of the good. Secondly, the more
heterogeneous and positively skewed the distri-
bution of potential contribution sizes, the more
likely group size will have a positive effect on
the provision of the good.

These two relations interact. Group size has a
positive effect whenever the interest and re-
source distributions are skewed enough relative
to the steepness of the cost function that the
effect of enlarging the pool of potential
contributors compensates for increased costs. If
costs increase only slightly with group size,
almost any heterogeneity in contribution levels
is enough to make larger groups more successful
than smaller ones. If costs increase substantially
with group size, however, then larger groups
will be less successful unless they are very
heterogeneous.[6]

THE PARADOX OF GROUP SIZE AND
THE NUMBER OF CONTRIBUTORS

In general, the complex interactions described
above are a difficult basis on which to build
useful substantive principles. However, they
have allowed us to recognize at least one
important paradoxical relation that has not
previously been appreciated. *When groups are
heterogeneous, a larger interest group can have
a smaller "critical mass."* That is, when a good
has high jointness of supply, it may be provided
by fewer people in a larger group than in a
smaller group.

There are precursors of this result in the
literature. Both Olson, briefly (1965, p. 29),
and Hardin, much more extensively (1982, pp.
67–89), argue that group heterogeneity has a
positive effect on the prospects for collective
action. Hardin gives several examples to show

---

[6] Using simulations, we have explored these numerical
relations extensively. However, we do not think that
numerical examples would do much to clarify the
substantive meaning of this result. The principles are as
we state them in the text, but real world situations differ
greatly in the actual forms of the two functions and, thus,
differ greatly in the outcome predicted by their
interaction.

that the especially interested and resourceful members of an interest group may provide collective goods that benefit many others. He proves that what he calls the "efficacious subgroup" (what we call the critical mass) will be smaller in a more heterogeneous group. We go one step farther and demonstrate that, if groups are heterogeneous, the critical mass will be smaller in a larger interest group.

We may illustrate the paradox with an example. Suppose the users of a computer facility are asked to chip in to buy a $125 software package that will be publicly available. For the moment, ignore the social process problem of coordinating contributions. It happens that the average person is willing to contribute $5. If the group is homogeneous, it takes 25 people to provide the good, a result that is invariant with group size.

In contrast, consider three installations with 100, 1,000, and 10,000 users, in which the distribution of resources among users is *heterogeneous*. Table 1 shows the proportions and actual numbers of users within each of these installations (groups) expected to have each whole number of resource units for two arbitrary distributions. The first is a normal distribution (which is, of course, symmetric), and the second is a moderately skewed lognormal distribution; both have a mean of $5 and a standard deviation of $1.

The data in Table 1 show that, regardless of group size, the simple fact of heterogeneity (around the same mean) reduces the minimum size of the critical mass. Even in the symmetric distribution, the smallest heterogeneous group (100), contains a minimum critical mass of size 20 (1 person contributing $8, 6 contributing $7, and 13 contributing $6), 5 less than the 25 contributors needed under homogeneity. This

Table 1. Computation of Critical Mass for Two Distributions (Normal and Lognormal with Mean 5 and Standard Deviation 1) and Three Group Sizes (100, 1,000, and 10,000)

A. Expected Numbers of Individuals Willing to Make Each Size Contribution. Rounded to Integers.

		Normal				Lognormal		
Value	Prob.	E(100)	E(1,000)	E(10,000)	Prob.	E(100)	E(1,000)	E(10,000)
0	.0000	0	0	0	.0000	0	0	0
1	.0002	0	0	2	.0046	0	5	46
2	.0060	1	6	60	.0747	7	75	747
3	.0606	6	61	606	.1877	19	188	1,877
4	.2417	24	242	2,417	.2197	22	220	2,197
5	.3829	38	383	3,829	.1821	18	182	1,821
6	.2417	24	242	2,417	.1273	13	127	1,273
7	.0606	6	61	606	.0815	8	81	815
8	.0060	1	6	60	.0497	5	50	497
9	.0002	0	0	2	.0296	3	30	296
10	.0000	0	0	0	.0175	2	17	175
11					.0103	1	10	103
12					.0061	1	6	61
13					.0036	0	4	36
14					.0022	0	2	22
15					.0013	0	1	13
16					.0008	0	1	8
17					.0005	0	1	5
18					.0003	0	0	3
19					.0002	0	0	2
20					.0001	0	0	1
21					.00008	0	0	1
22					.00005	0	0	1
23					.00003	0	0	0
24					.00002	0	0	0
25					.00001	0	0	0

B. Computation of Size of Critical Mass.

Distribution	Group Size	Size of Critical Mass	Detail
Normal	100	20	1 @ 8; 6 @ 7; 12 @ 6; 1 @ 3.
Normal	1,000	17	6 @ 8; 11 @ 7.
Normal	10,000	16	2 @ 9; 13 @ 8; 1 @ 3.
Lognormal	100	15	1 @ 12; 1 @ 11; 2 @ 10; 3 @ 9; 5 @ 8; 2 @ 7; 1 @ 1.
Lognormal	1,000	9	1 @ 17; 1 @ 16; 1 @ 15; 2 @ 14; 3 @ 13; 1 @ 10.
Lognormal	10,000	7	1 @ 22; 1 @ 21; 1 @ 20; 2 @ 19; 1 @ 18; 1 @ 6.

minimum number declines slightly with the group size: to 17 for the installation with 1,000 users, and to 16 for the largest group of 10,000 members.

This pattern is more pronounced when the resource distribution is more skewed. For the moderately skewed distribution in this example, the minimum number of contributors is 15 for the smallest group, 9 for the medium-sized group, and 7 for the largest group. Extremely skewed distributions, in which some members might be willing and able to contribute 100 times more than the mean of all others, would show even more pronounced effects, so that one or two people might be able to provide the good for the whole group.

It should be clear that this pattern is not dependent on any particular distribution, but rather may arise whenever a group is heterogeneous in the sizes of its members' potential contributions. The mechanism causing the paradox is really very simple: the expected number of individuals willing and able to give at any specific contribution level will always be higher for a larger group. Since collective goods with pure jointness of supply have a fixed cost that does not vary with the size of the group enjoying the good, the greater expected number of large contributors in a larger group means that, in general, fewer people will be needed to achieve a given *total contribution* size than in a small group.

GROUP SIZE AND THE SOCIAL PROCESSES FOR COLLECTIVE ACTION

Olson is right: there are many public goods which will never be provided by individuals acting in independent isolation. However, Olson argues that, even if we allow for social processes, the group size effect would obtain, since such social processes, as well as feelings of group solidarity, are more likely to overcome the collective dilemma in "moderate" sized groups than in large ones (p. 48). This argument is seriously flawed by a floating conception of what the "group" is. When a good has jointness of supply, it is irrelevant to those who contribute how many others there are "out there" in the interest group who might benefit. When a "social" solution to the collective dilemma is required, what matters is the relationship among the possible contributors in the critical mass, not the relationship among everyone in the interest group. Paradoxically, the size of the critical mass will be smaller when the size of the interest group is larger, and social processes may be *more* beneficial in larger interest groups.

Because larger interest groups have more total resources, they are generally more likely to have the *possibility* for a successful collective action.

Especially when goods have high and "lumpy" costs (i.e., where a large minimum amount is needed to provide any of the good, such as a bridge), smaller interest groups may be simply unable to supply enough resources, no matter how well they organize. Where a larger group might need to mobilize only 5 percent of its potentially available resources to provide a good, the smaller group might require 100 percent of its resources, or more.

There are doubtless some small interest groups with the kind of social structure that would permit them to mobilize 100 percent of their members to action, and it is likely to be very exciting when it happens. But it is probably more common to see a critical mass coalesce within a larger interest group. There are costs to organizing and coordinating contributions by a number of people, and those costs are usually higher the more contributors there are involved. Thus, it will generally be much easier and cheaper to organize a collective action involving a small number of contributors from within a large interest group than one involving a larger number of contributors from within a small interest group.

Of course, it may be especially difficult and costly for the small number of potential contributors in a very large interest group to find one another and coordinate their actions. If society were organized randomly, this would always be a serious problem for collective actors. If social ties were distributed randomly across a large city, it would be unlikely that the five people who would be willing to contribute $1 million each to a geology museum, or the ten people who would be willing to devote six months of their lives to organizing a nuclear freeze campaign, would ever meet. In fact, the real world surely contains many "interests" whose distributions are essentially random, and about which collective action is very unlikely. But randomness is *not* the rule. Especially wealthy people know most of the other especially wealthy people. Potential political activists associate themselves with events and organizations expressing their political concerns. City residents who would be most harmed by a proposed expressway live near each other, in its path.

The problem of collective action is not whether it is possible to mobilize every single person who would be benefited by a collective good. It is not whether it is possible to mobilize everyone who would be willing to be mobilized. It is not even whether all the members of some organization or social network can be mobilized. Rather, the issue is whether there is some social mechanism that connects enough people who have the appropriate interests and resources so that they can act. It is whether there is an

organization or social network that has a *subset* of individuals who are interested and resourceful enough to provide the good when they act in concert, and whether they have sufficient social organization among themselves to act together.

In one sense, our argument is that Olson's "large group" problem is resolved by the "small group" solution. Olson is right in saying that collective action almost never takes the form of small, unnoticeable contributions from thousands or millions of isolated individuals. If everybody's interest or resources are equally small, collective action will generally not happen, no matter how big or small the interest group. Collective action arises around those interests for which there are a group of especially interested and resourceful individuals who are socially connected to one another. (For a much fuller analysis of the effects of social ties within groups, see Marwell, Oliver, and Prahl, forthcoming.)

The small number of wealthy people are able to act collectively to get what they want not because there are few of them, but because they are wealthy. Resources and interests being equal, movements on behalf of very large constituencies often are more successful than movements on behalf of tiny minorities. Large interest groups do sustain more collective action then smaller ones, when costs are equal and the individuals in the groups have comparable interest and resource levels. Resources and social organization are the problem, *not* group size. If a group is heterogeneous enough that it contains a critical mass who can make large contributions, and if those members are socially connected to one another so that they can act in concert, collective action *is* possible and *more* likely in larger groups.

Our theoretical analysis is consistent with much recent empirical scholarship on social movements. It is never the case that all women (Freeman 1983), all blacks (Morris 1981, 1984), everyone opposed to the reopening of the Three Mile Island reactor (Walsh and Warland 1983), everyone for a clean environment (Mitchell 1979), or all northern whites concerned about voting rights in the South (McAdam 1986) are mobilized, nor is the existence of a large mass of "free riders" any particular hindrance to the mobilization or success of a movement. In fact, public opinion polls identifying large pools of nonactivist adherents to a cause tend to help the cause, not hurt it. What matters for successful mobilization is that there be enough people who *are* willing to participate and who are also reachable through social-influence networks. Empirical accounts of actual social movements and movement organizations show over and over that most of the

action originates from a relatively small number of extremely active participants.

The "free rider" dilemma, correctly analyzed, is the problem of not being able to make a big enough difference in the outcome to compensate for the costs one bears. Thus understood, the theory of collective action does not predict that collective action will *never* occur, but rather that it will not take the form of small isolated contributions. Instead, the theory of collective action explains why most action comes from a relatively small number of participants who make such big contributions to the cause that they know (or think they know) they can "make a difference." In social movements, these contributions are usually time and energy, not money.

Theory and empirical research also agree in tending to discount the causal significance of the size of the aggrieved population as a direct determinant of collective action. Current research stresses the importance of social networks and organizational resources among some interested subset of the population, coupled with "political opportunities" created through party politics (e.g. McAdam 1982; Jenkins 1987). Full-time professional activists (McCarthy and Zald 1973, 1977) are also seen as important, although less so than several years ago.

Let us not, however, conclude that the masses are irrelevant for collective action. We have shown theoretically that larger groups *should be* more likely to give rise to collective action than smaller ones (given the jointness of supply of so many collective goods), and it is empirically true that very large social movements tend to arise from very large mass bases. However, undifferentiated impoverished masses do not usually support social movements. What seems to be critical is the presence of a *minority of the aggrieved population* who are well educated or especially politically conscious, who have high discretionary time, or who are economically independent of the oppressors.[7] Larger populations are likely to have larger numbers of these "unusual" members, and the size of their potential contributions is likely to be larger.

The more obvious effect of interest-group size is also important: larger populations generally have more total resources than smaller ones. This has frequently been ignored by those who are theoretically sophisticated, because it is understood that those resources do not automat--

---

[7] Perhaps we should cite Lenin on this point, as well as the social movements literature. We should stress that we are emphasizing the theoretical importance of differentiation *within* the aggrieved population, which is very different from an "external resources" argument, which has fared badly empirically (e.g. McAdam 1982; Jenkins and Eckert 1986).

ically or easily become contributions. But one thing the small critical mass of large contributors can do is invest time, energy, and money in organizing and coordinating events that draw in and make use of small contributions. The critical mass can use preexisting organizations and networks to create the social conditions under which small contributors will participate in a march or demonstration. They can pay the overhead for large mass mailings to solicit small monetary contributions under circumstances that make the donors' costs low relative to their psychic "benefits" (see Oliver and Furman forthcoming for more discussion). The larger the total size of the interest group, the larger the potential gain from either of these strategies.

As a final note, it is important that our argument not be read only as a critique of Olson. Instead, we have tried to constructively describe *how* group size affects the prospects for collective action. Of course, the kind of cost/benefit considerations we analyze are not all that are involved in collective action. But if we understand how the cost structure of a good, and the distribution of resources and interest across the pool of people interested in that good, interact with the size of the group to produce structural constraints on the possibilities for action, we can use this information as a baseline for investigating the effects of other factors on the prospects for collective action. The time is long past for sociologists to stop debating whether free riding does or does not occur (sometimes it does and sometimes it does not) and get on with specifying the conditions that favor or hinder collective action.

REFERENCES

Chamberlin, John, 1974. "Provision of Collective Goods as a Function of Group Size." *American Political Science Review* 68:707–16.

Freeman, Jo. 1975. *The Politics of Women's Liberation.* New York: Longman.

Frohlich, Norman and Joe A. Oppenheimer. 1970. "I Get By With A Little Help From My Friends." *World Politics* 23:104–20.

Hardin, Russell. 1982. *Collective Action.* Baltimore: Johns Hopkins University Press.

Head, John G. 1974. *Public Goods and Public Welfare.* Durham, NC: Duke University Press.

Jenkins, J. Craig. 1987. "Interpreting the Stormy Sixties: Three Theories in Search of a Political Age." *Research in Political Sociology* 3:269–303.

Jenkins, J. Craig and Craig M. Eckert. 1986. "Channeling Black Insurgency: Elite Patronage and Professional Social Movement Organizations in the Development of the Black Movement." *American Sociological Review* 51:812–29.

Latane, Bibb and Steve Nida. 1981. "Ten Years of Research on Group Size and Helping." *Psychological Bulletin* 89:308–24.

Marwell, Gerald. 1970. "Comment on Scott and El-Assal." *American Sociological Review* 35:916.

Marwell, Gerald, Pamela E. Oliver, and Ralph Prahl. Forthcoming. "A Theory of the Critical Mass. III. Social Networks and Collective Action." *American Journal of Sociology.*

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930–1970.* Chicago: University of Chicago Press.

———. 1986. "Recruitment to High-Risk Activism: The Case of Freedom Summer." *American Journal of Sociology* 92:64–90.

McCarthy, John and Mayer Zald. 1973. *The Trend of Social Movements in America: Professionalization and Resource Mobilization.* Morristown, NJ: General Learning Press.

———. 1977. "Resource Mobilization in Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212–39.

McGuire, Martin C. 1974. "Group Size, Group Homogeneity, and the Aggregate Provision of a Pure Public Good Under Cournot Behavior." *Public Choice* 18:107–26.

Mitchell, Robert Cameron. 1979. "National Environmental Lobbies and the Apparent Illogic of Collective Action." Pp. 87–121 in *Collective Decision Making: Applications from Public Choice Theory,* edited by Clifford S. Russell. Baltimore: Johns Hopkins University Press for Resources for the Future.

Morris, Aldon. 1981. "Black Southern Student Sit-In Movement: An Analysis of Internal Organization. *American Sociological Review* 46 (December):744–67.

———. 1984. *The Origins of the Civil Rights Movement: Black Communities Organizing for Change.* New York: The Free Press.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356–75.

Oliver, Pamela and Mark Furman. Forthcoming. "Contradictions Between National and Local Organizational Strength: The Case of the John Birch Society." In *Organizing For Change,* edited by Bert Klandermans. Greenwich, CT: JAI Press.

Oliver, Pamela, Gerald Marwell, and Ruy Teixeira. 1985. "A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Goods." *American Journal of Sociology* 91:522–56.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Piliavin, Jane Allyn, John Dovidia, Sam Gaertner, and Russell Clark, III. 1981. *Emergency Intervention.* New York: Academic Press.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36:387–89.

Scott, J.W. and M. El-Assal. 1969. "Multiversity, University Size, University Quality and Student Protest—An Empirical Study." *American Sociological Review* 34:702–9.

Spilerman, Seymour. 1970. "The Causes of Racial Disturbances: A Comparison of Alternative Explanations." *American Sociological Review* 35:627–49.

Walsh, Edward and Rex Warland. 1983. "Social Movement Involvement in the Wake of a Nuclear Accident: Activists and Free Riders in the TMI Area." *American Sociological Review* 48:764–80.

ACTION AND INFORMATION IN THE JOB MOBILITY PROCESS: THE SEARCH DECISION*

CHARLES N. HALABY
University of Wisconsin-Madison

Current sociological models of job mobility suppress the intervening action and information mechanisms that link structural constraints to job shifts. To remedy this, this paper joins core features of structural-mobility models to key elements of economic approaches to job change in an analysis of the employed worker's decision to search the market for a better job. The analysis highlights the complementarity of structural and action approaches by conceptualizing the forces governing the search decision in terms of the principal dimensions—opportunities and achievement gains—underlying job-mobility models. The basic framework is extended to include the quality of a worker's initial screen of a prospective job as a mechanism that promotes search. The concepts of opportunities, gains, and screen quality are then used to explain male-female differences in mobility orientations. The empirical findings both reinforce structuralist assumptions and underscore the promise of applying to the study of mobility processes the kind of action approach recently advocated by Coleman (1986).

INTRODUCTION

Sociological models of voluntary job mobility have little to say about worker action and information as mechanisms of change in achievement. The dominant theoretical framework finds structural constraints joining with worker attributes and productive resources to yield job shifts without any explicitly modelled worker intervention. Although sociologists understand that structural constraints are mapped into job shifts by an intervening process of choice and action, this process is usually relegated to the background and invoked only informally to lend intuitive appeal to the structural explanation. The information that workers have about their job and labor market position, as well as their efforts to search for better jobs, are *formally* irrelevant because mobility is treated, not as a voluntary decision or action, but as a structurally induced event. In bypassing the process by which workers evaluate their market prospects and actively construct their careers, the structural approach renders the mobility *decision* analytically unproblematic.

These features are evident in Sørensen's (1977, 1979) vacancy-competition model, the sharpest structural formulation of job mobility in

the sociological literature. Together with Tuma (1976), Sørensen established the basic analytical framework underlying much of the job-mobility research of the last decade (Felmlee 1982; DiPrete 1981; Carroll and Mayer 1986; Waite and Berryman 1986). In his model, job shifts result solely from the interaction of exogenous structural opportunities and worker productive resources. While Sørensen, like other structuralists, recognizes the necessity of an intervening choice mechanism—workers attempt to maximize status (1979, p. 368; 1977, p. 970)—such a mechanism is not integral to the model. On the contrary, the model suppresses the process by which workers assess their mobility chances and then choose to act accordingly.

Coleman (1986) has recently chastised sociological researchers for just this kind of practice: causally relating "structure" to outcomes without a theory of the intervening action process. His criticism applies with special force to job-mobility research because sociological efforts unfold alongside the action- and information-based models of economics. Over the last two decades, the marriage of human capital and information theory has produced an array of economic models that provide a rich theoretical portrait of the decision calculus surrounding job mobility (Stigler 1962; Parsons 1973, 1977; Jovanovic 1979b, 1984). Imperfect information is a key feature of the new models, not only because it implies that more information may lead workers to better jobs, but because it renders the knowledge and perceptions of workers an analytically significant part of the job-mobility process. Indeed, recent work expands the notion of mobility from a discrete event of limited duration to an unfolding course

of action (McCall 1970; Lippman and McCall 1976; Burdett 1978). Under this conceptualization, the decision to search for and accept alternative job offers itself becomes a central and problematic feature of the mobility process.

This paper joins key elements of economic treatments of search and information to core features of the structuralist model of job mobility. The aim is to incorporate workers' knowledge and perceptions of the labor market into the sociology of voluntary job change. This should not be construed as an attempt to replace the structuralist model with a choice-based alternative. On the contrary, the core concepts of the structuralist approach are an effective framework for a choice-based model of worker mobility orientations. To this end, I will show that the forces that structural models reveal as governing mobility find expression in the decision calculus surrounding job change.

The vehicle for the analysis is the decision of employed workers to search the marketplace for wage offers. This is a timely and reasonable starting place for highlighting the complementarity of structural and action approaches. It is timely because contemporary economic research on the employed worker's decision to search is barely underway and limited to demographically narrow segments of the labor force (Black 1981; Kahn and Low 1984). It is reasonable because search not only signals workers' impending mobility but is itself a productive mobility strategy. Employed searchers are three times more likely to change jobs than comparable nonsearchers (Black 1981), and the economic returns to job change are higher for searchers than for comparable workers who shift without searching (Black 1980; Kahn and Low 1982). Hence, the propensity to search is linked directly to mobility decision making, and thereby reveals the forces shaping worker efforts to secure mobility gains.

In light of this, the first task is to link conceptually the major dimensions of the search-decision process to the major dimensions of the shift process that has been the subject of structural models. I do this with a model that assumes that the theoretical mechanisms at the core of Sørensen's theory—job opportunities or vacancies, and exploitable achievement (e.g., status or wage) gains—are also the principal dimensions of the calculus governing the employed worker's decision to search. This provides a basis for using the analysis of search to elucidate several aspects of the mobility-action process that in the past have been either inadequately treated or suppressed altogether.

First, the analysis clarifies the critical role of job opportunities and potential status or wage gains in the mobility process. The treatment of these theoretical constructs in empirical studies of job shifts has been problematic because it demands measures of hard-to-observe "objective" structural features of workers' market position. The analysis of search offers a way out of this dilemma because the decision to look for a better job may, in the presence of imperfect information and bounded rationality constraints, be plausibly framed in terms of workers' *information about* and *subjective estimates of* opportunities and gains.

Second, the analysis addresses a rarely examined alternative conceptualization of how information shapes mobility. In Sørensen's model, for example, a job shift may be viewed as a worker's reaction to the arrival of new information about an *alternative job* opening. A second type of model posits the arrival of new information about the worker's *current job* as the mechanism that triggers mobility activity (Jovanovic 1979a, 1979b; Lippman and McCall 1981). This model assumes that the value of certain job properties and working conditions cannot be accurately *screened* by the worker prior to employment. Rather, such information is acquired through actually working in the job. The implication is that the quality of the initial screen bears on a worker's risk of learning by experience that he has formed a bad match, which in turn may lead to regret, search, and, eventually, mobility. This analysis assesses the significance of an imperfect initial screen as a mechanism promoting search.

Finally, opportunities, gains, and screen quality have all figured prominently in discussions of male-female differences in the rate of mobility within the labor market. Yet, the explanatory power of these factors is still largely speculative, and even less is known about their implications for sex differences in search propensity. The available studies of search by employed workers have been limited to male samples (Black 1981) and have been otherwise unsuitable for gauging the implications of sex differences in market perceptions and information for the decision to search for a better job. An important unresolved issue is how the experience of sexual discrimination shapes the mobility orientations of women workers.

THEORETICAL FRAMEWORK

Search, Opportunity, and Gain

A search model formulated by Parsons (1973) serves as a framework for the theoretical and empirical analysis. The model assumes that a worker's decision to search involves a single property of jobs—the wage rate—and that there exists a stable distribution of wage offers W_a from potential employers with vacancies. The worker knows the form and parameters of the wage offer distribution, but not the value of particular wage bids he may elicit if he searches. To be of value, a wage bid w_a resulting from

search must exceed the *critical wage* $W_c = W_o + k$, where W_o and k are current wage and transfer costs, respectively. The decision to search is governed by the expected economic returns, which will depend on (1) the probability of locating a job vacancy that offers a wage $w_a > W_c$; and (2) the expected value of the wage gain $(W_a - W_c)$. When these two factors are combined, the expected returns to search is given by

$$R = q(v)[\int_{W_c}^{\infty} (W_a - W_c)f(W_a)dw_a] \qquad (1)$$

where f is the density of the wage-offer distribution. The worker's estimate of the probability of eliciting an offer if he searches is given by a function $q(v)$ of job vacancies v $(\partial q(v)/\partial v > 0)$.

This model states that the returns to search, and, hence, the propensity to search, is a function of job opportunities, $q(v)$, and exploitable market wage gains, in brackets. In other words, the forces governing search in this model are conceptually identical to those governing shifts in Sørensen's model. Indeed, even the forms of the two models—opportunity and gain interact to determine the outcome—are the same (see Sørensen 1977, eq. 19; 1979, eq. 6).

The model of eq. 1 may be usefully extended by assuming that the wage-offer distribution is normal with mean μ_a and variance σ^2. Then, integrating the right-hand side yields (Parsons 1973),

$$R = q(v)[\sigma^2 f(W_c) \\ - (W_c - \mu_a)\phi(W_c)] \qquad (2)$$

where $\phi(W_c) = 1 - F(W_c)$, and f and F are the normal density and distribution functions. This expression shows that the propensity to search will (1) increase as labor demand $(\partial R/\partial q(v) > 0)$ and the mean and variance of the wage-offer distribution increase $(\partial R/\partial \mu_a, \partial R/\partial \sigma^2 > 0)$; and (2) decrease as the worker's current wage, transfer costs, and, hence, critical wage increase $(\partial R/\partial W_c = \partial R/\partial W_o = \partial R/\partial k < 0)$.

Eq. 2 figures prominently in the ensuing discussion and analysis because it concisely represents the major theoretical constructs informing much sociological and economic research on voluntary job mobility. The next two sections use this form of the model to review the treatment of opportunity and gain in past studies, and suggest alternative strategies suitable for the analysis of search propensity.

Vacancies, Offers, and Search

Although national and industry studies show that aggregate quit rates are highly responsive over time to changes in the job-vacancy rate (Barron and McCafferty 1977; Parsons 1973), in

individual-level studies labor demand variables have yielded weak results. Economic studies commonly include the size, regional location, and tightness of the local labor market as indicators of $q(v)$ (Black 1981; Blau and Kahn 1981), but without much success. A recent sociological study by Waite and Berryman (1986) included similar measures of market conditions, but with the same lack of success. This may explain why most sociological analyses in the spirit of Tuma and Sørensen do not even bother to take explicit account of labor demand (Tuma 1976; Sørensen and Tuma 1981; Felmlee 1982; Carroll and Mayer 1986). Many studies rely instead on measures of worker resources (e.g., education and ability), claiming that these tap "access to opportunities." This practice is ambiguous at best, since "resources" also serve, as discussed below, as implicit controls for the mean μ_a of the wage-offer distribution.

The absence of effects for job opportunities in these studies is understandable. While aggregate quit rates respond to aggregate vacancy rates, the latter, as indicated by conventional measures of $q(v)$, can only crudely tap the demand forces shaping the search and quit decisions of individual workers. In the first place, the mobility activities of individual workers may follow more skill-specific labor-demand functions. More importantly, the propensity to search should respond, not to the aggregate volume of vacancies alone, but to vacancies with *acceptable wage offers*. This can be seen by rewriting eq. 2 as

$$R = \sigma^2 Q_1 - (W_c - \mu_a)Q_2 \qquad (3)$$

where $Q_1 = q(v)f(W_c)$ and $Q_2 = q(v)\phi(W_c)$. This equation breaks up the contribution of $q(v)$ into two parts: Q_1, which measures the likelihood the worker will elicit a minimally acceptable offer, that is, an offer in the vicinity of the critical wage W_c; and Q_2, which measures the likelihood the worker will elicit an offer exceeding the critical wage. While it would be difficult to obtain "objective" measures of either Q_1 or Q_2, what may count most in the decision to search is workers' *information about* job opportunities paying an acceptable wage for skills of the kind they can supply. Hence, as an alternative to the usual aggregate proxies for $q(v)$, the role of opportunity or labor demand in the worker's decision calculus may be represented by subjective estimates of either Q_1 or Q_2. Since the term involving Q_1 makes a larger contribution to search than the term involving Q_2, a measure of Q_1 will be used here.[1]

---

[1] In equation 2, $[\sigma^2 f(W_c)] - (W_c - \mu_a)(1 - F(W_c)] > 0$ when $\sigma > 0$, so that $\sigma^2 Q_1 > (W_c - \mu_a)Q_2$.

Wage Gains: Past Practice and an Alternative

Efforts to tap how much a worker stands to gain by changing jobs have understandably focused on measuring $(W_c - \mu_a)$ in eq. 2. Sociological efforts have been implicit: the workers' critical wage W_c as indicated by current wage W_c is augmented by measures of productive resources that act as controls for the mean μ_a of the wage-offer distribution (although see Tuma 1976, and Carroll and Mayer 1986, for studies in which no measure of current wage is included). Economic studies frequently take a more explicit approach: the critical wage W_c is set equal to W_o and then expressed as a deviation from an estimate of μ_a generated by an auxiliary earnings function (Black 1981; Viscusi 1980). Both approaches assume that workers infer the potential wage gains from a job change by comparing directly their own wage to a known market wage-offer mean.

A different conceptualization of how workers gauge that wage gains may accompany a job shift, and that search is therefore warranted, yields an alternative means of imputing wage gains as subjectively known to workers. Instead of comparing their own wage to a market mean, assume workers initially estimate their wage-gain potential based on their knowledge of the causal mechanism responsible for generating a *distribution* of wage offers across employers for a worker with a fixed endowment of productive resources. If workers know why different employers offer different wages to a given worker, then they can infer the amount by which a typical alternative offer might be expected to exceed their current wage.

One mechanism accounting for the wage-offer distribution is variation in the extent to which employers exploit (i.e., employ) a given worker's skills and abilities. Such variation would induce variation in a worker's productivity across employers, and thus yield an alternative wage-offer distribution. This mechanism, derived from economic theory (Jovanovic, 1979b), is consistent with structuralist mobility models. In Sørensen's model, for example, a worker is impeded by the opportunity structure from immediately moving into a job paying his maximum equilibrium lifetime wage W_m; in the meantime, he takes a job paying $W_o < W_m$. This implies that the worker's productivity in the job paying W_o is less than it would be in the highest attainable job paying W_m. But since a worker's endowment of productive resources is, under the model, fixed over time, the disparity in productivity and, hence, wages for the two jobs must have a different source. One possibility is that the job paying W_o does not as fully exploit the worker's productive resources as the job paying W_m. Hence, job shifts in Sørensen's

model may be understood as moves by workers into jobs that make better use of their skills and abilities, and, therefore, pay more.

This conceptualization suggests an alternative means by which workers may estimate the potential wage returns to search, namely, by assessing the rate, say ρ $(0 < \rho < 1)$, at which the current job exploits their productive resources. That is, workers may infer their wage gains by assessing their productivity on the current job *relative* to what it would be on a job that made better use of their resources. Workers who believe their skills are fully employed on the current job will see little chance of boosting their productivity through a job change, will arrive at pessimistic estimates of their exploitable market wage gains, and will be unlikely to search. Conversely, workers who believe they possess an abundance of unexploited skills will arrive at more optimistic estimates of the potential returns to search. One advantage of this conceptualization is that it demands of workers only information of which they have first-hand knowledge. The worker's personal knowledge of what shall hereafter be called his *relative productivity* ρ on the current job, plus his knowledge that $max[\rho] = 1$, provide a baseline for assessing the possible productivity/wage gains from job change. The empirical analysis will compare the explanatory power of this approach to the usual practice of using imputed money measures of $(W_c - \mu_a)$.

EXTENSIONS

The Quality of the Screen

The information workers hold about the value of a prospective job is highly imperfect because certain properties of jobs—what economists call *experience properties*—are difficult to assess in the absence of actual employment (Jovanovic 1979a; McCall and Lippman 1981). Moreover, because of differences among employers in the signals sent to prospective employees, or differences among workers in their informal ties to the workplace (Bridges and Villemez 1986; Granovetter 1974), workers vary in the quality of their initial *screen*, that is, in the accuracy with which they anticipate and appraise job properties prior to employment. The initial screen is an important mechanism matching workers with heterogeneous skills and tastes to jobs with heterogeneous requirements and working conditions. A poor screen increases the risk of a mismatch; once on the job, the worker may learn that it is less valuable than anticipated because it fails to match her skills and tastes. The arrival of this new information may lead workers to revise downward their assessment of the job's value; search may ensue.

Sociologists have not paid much attention to

the quality of the screen as a mechanism generating mismatches and, hence, mobility. Yet recent findings (Datcher 1983) and informed speculation indicate that screen quality may be a significant factor underlying observed differences in worker search and quit propensities. Some have suggested that the higher overall rate of quitting among women compared to men may be due to poor quality screens that expose women to an especially high risk of mismatch (Viscusi 1980; Meitzen 1986). A similar explanation might apply to the well-known difference in rates of turnover between core and periphery sectors: workers recruited to the core may have better information about prospective jobs than those recruited to the periphery.[2] Hence, mobility differences that exist across types of workers or markets, and that are now commonly attributed to "good" versus "bad" jobs, may be due partly to differences in screen quality.

More crucial for our purposes is how the quality of the initial screen may be implicated in the formation of worker assessments of their relative productivity on their current job. If workers know *ex ante* (i.e., by *inspection*) the extent to which a prospective job will use their productive resources, then the quality of the screen is clearly irrelevant. But it is more likely that relative productivity is itself subject to an imperfect screen: the extent to which a job employs a worker's skills and abilities may not be known prior to employment, but rather may be revealed only through working (Jovanovic 1979a). In this case, the quality of the screen would be implicated in the estimation of the effects of relative productivity on search: an initial screen that affects search by increasing the risk of mismatch on productivity attributes may also affect search by increasing the risk of mismatch on other job properties. The effects of relative productivity will be confounded with the effects of other, unmeasured types of mismatches. Such confounding may be reduced by controlling for variation among workers in screen quality.

Male-Female Differences in Search

Previous studies of employed search have been based on male workers (Black 1981; Kahn and Low 1984), so the reasoning about sex differences in search propensity relies on research on quitting. The available evidence indicates that male rates of quitting to move within the labor

market—the only type of moves that are germane—exceed female rates (Barnes and Jones 1974). Taking this generalization as a point of departure, I examine the alternative search mechanisms that may give rise to it.

The usual explanation for the lower female quit rate is that market discrimination and segregation limit the job openings available to women and the wage gains they can hope to achieve (Barnes and Jones 1974; Stoikov and Raimon 1968). Three alternative patterns of sex differences in search propensity are consistent with this reasoning. First, alternative market opportunities are better for men, both men and women are accurately informed about this, and, therefore, men search more and eventually quit more than women. Second, labor market opportunities are better for men, but women, because of their lower labor-force experience, are misinformed in such a way that their perceived opportunities are approximately equal to those of men. Under this assumption men and women might be equally likely to *embark* on search, although, presumably, women would stop searching sooner and eventually quit less as they learned of their true, more limited opportunities. Third, women may have a higher propensity to search, even though discrimination/segregation so limits their opportunities that they ultimately quit less than men. The argument is that, because of their more intermittent labor-force participation, women have less accurate information prior to employment about the "experience" properties of prospective jobs, which ultimately exposes them to a higher risk of mismatch. In other words, the quality of the initial screen is lower for women than for men (Viscusi 1980; Meitzen 1986).

A narrower version of this last hypothesis is of some interest. It argues that women are exposed to the risk of a type of highly unfavorable experience to which men are largely immune—sexual discrimination (Viscusi 1980; Meitzen 1986). Even if men and women were equally well-informed prior to employment about all other properties of a prospective job, sexual discrimination alone would put women at risk of a bad match and increase their search propensity. This hypothesis assumes that felt sexual discrimination is difficult to accurately anticipate, not only because of veiled signals from employers, but also because it is not (perceived by women to be) pervasive. If women could with certainty anticipate discrimination from virtually all employers, then the actual on-the-job experience of discrimination would not constitute *new information,* and therefore would not lead women to revise downward their assessment of the job's value. Furthermore, pervasive discrimination would preclude job change as a means of avoidance.

---

[2] This may be due to a number of factors: the variance of experience properties may be less in the core than in the periphery; core employers may send better signals; core workers may be more aggressive in seeking job information.

But if felt discrimination cannot be anticipated with certainty, then the on-the-job experience of discrimination would constitute new information and might lead women to lower their assessment of the current job's value and to search for alternative employment. These conjectures are of interest because of speculation about sexual discrimination "pushing" women into quitting. Osterman (1982) has addressed this issue indirectly with evidence showing that affirmative action programs reduce female quit rates. The analysis below sheds further light on this subject.

DATA AND MEASUREMENT

The data come from the 1977 Quality of Employment Survey (Quinn and Stanley 1979). This survey used a national probability sample ($N = 1,515$) of persons 16 years or older who were working for pay 20 hours or more per week. The analysis here is based on the unweighted sample of full-time (≥ 30 hours per week) workers under 65 years of age who "work for someone else" ($N = 1,204$).[3]

The dependent variable indicates the respondent's propensity to search for a job with another employer. The question reads: "Taking everything into consideration, how likely are you to make a genuine effort to find a new job with another employer within the next year?" ("very likely" = 5, "somewhat likely" = 3, and "not at all likely" = 1).[4]

The independent variables, described in Table 1, divide into two groups: (1) those indicating "objective" features of a worker's job and market position as conventionally measured in sociological and economic treatments of search or job change; and (2) those indicating worker subjective assessments of market opportunities, relative productivity, and the quality of the initial screen. I begin by discussing the former measures.

The most common money measure of the potential wage gain from searching or shifting is the current wage component of the critical wage, here measured as the logarithm of weekly earnings (LNWAGE). An alternative measure is the negative of the residual from an "objective" market earnings function, namely WAGEGAIN = (LNWAGE—LNWAGE) (Viscusi 1980).[5] Finally, Black (1981) has suggested that this measure be corrected for differences in marginal search costs by dividing by current wage. This yields %GAIN = WAGEGAIN/LNWAGE. Note that WAGEGAIN and %GAIN measure wage gains directly, while LNWAGE does so inversely.

Other variables pertaining to the net economic returns to search are LOSEJOB, BENEFITS, UNION, SUPERVISOR, TENURE, AGE, and LFEXP. Since LOSEJOB is an inverse measure of the future economic returns to the current job, an increase in LOSEJOB is like a decline in current wage and should increase search. Conversely, fringe benefits (BENEFITS) and unmeasured pecuniary rewards associated with managerial status (SUPERVISOR) increase the critical wage and should decrease search. UNION may also capture such pecuniary rewards, as well as nonpecuniary compensation in the form of grievance procedures and other legal protections against managerial arbitrariness (Farber and Saks 1980; Freeman 1980). TENURE is a proxy for the transfer costs associated with the loss in earnings tied to firm-specific skills and should decrease search; TENURE^2 captures the decline in the rate of investment in firm-specific skills. Labor force experience (LFEXP) is sometimes used by economists as a measure of the mean μ_a of the wage-offer distribution and by Sørensen as a measure of the difference between current wage and maximum lifetime wage. AGE captures the remaining length of work life and serves as an inverse measure of the long-run value of wage gains (Gronau 1971). AGE may also indicate job opportunities if employers find older workers less attractive because of the shorter time period over which to recoup fixed costs of hiring and training (Hutchens 1986).

Research shows that employer size and industry are associated with returns to the current job. Because size of employer is positively associated with current and future economic returns to the current job (Garen 1985), and with a host of other advantageous job conditions (Baron and Bielby 1984), FIRM-SIZE should decrease search. This would be consistent with the finding that size reduces the rate of quitting to change employers (Carroll and Mayer 1986). Since stable, continuous employ-

---

[3] In keeping with *ASR*'s proposed policy, I will make available as hard copy or via BITNET the means, standard deviations, and correlations needed to estimate all the models presented in this paper. The raw data for the *QES* has been archived and available for years from the Inter-University Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, MI 48106.

[4] This particular item has been used in other studies based on the *QES* as well as other surveys, and has consistently been shown to have high construct and criterion validity with respect to actual quits (Viscusi and O'Connor, 1984). It may be safely assumed, therefore, that its construct and criterion validity with respect to search is even higher.

[5] LNWAGE was generated by separate male and female regressions of LNWAGE on schooling, labor-force experience (and its square), four occupation dummies, three region dummies, three city size dummies, union status, supervisory status, race, hours worked per week, and a dummy for core industrial sector.

Table 1. Variable Names and Descriptions

Variable	Descriptions
LNWAGE	Natural logarithm of mean weekly earnings.
WAGEGAIN	Negative of the residual from a market earnings regression for LNWAGE (note 8).
%GAIN	WAGEGAIN/LNWAGE.
LOSEJOB	"How likely is it that during the next couple of years you will lose your present job and have to look for a job with another employer?" (1 = not at all likely; 4 = very likely)
BENEFITS	Scale constructed by counting the number of benefits the respondent receives on the current job. The benefit categories were retirment program, life insurance, savings plan, paid vacation, medical insurance, paid sick leave, dental and eyecare benefits, and, for women only, paid maternity leave and maternity leave with full re- employment rights.
TENURE	Years of service with employer.[1]
AGE	Respondent's age in years.
LFEXP	Respondent's years of labor-force experience.
SUPERVISOR	Dummy coded 1 if respondent supervises the work of others.
UNION	Dummy coded 1 if respondent's job is covered by collective bargaining agreement
FIRMSIZE	Number of people who work for employer at same location as respondent.[1]
CORE	Dummy coded 1 if respondent's job is in "core" sector as defined by the Beck, Horan, and Tolbert (1978) typology.
BIGCITY	Dummy coded 1 if respondent resides in the central city or suburbs of the 12 largests SMSAs.
WEST	Regional dummy coded 1 if respondent resides in the West.
FEMALE	Dummy coded 1 for female respondents.
NONWHITE	Dummy coded 1 for nonwhite respondents.
SCHOOLING	Respondent's years of schooling.
OTHERINCOME	Respondent's annual earnings expressed as percentage of total annual family income from all sources.
OTHERJOB	Dummy coded 1 if respondent works for pay on a job other than his main job.
MARRIAGE	Dummy variable coded 1 for married, spouse present.
KIDS<6	Number of children less than six years old in household unit who are respondent's or respondent's spouse's.
KIDS<18	Number of children age 6–17 in household unit who are respondent's or respondent's spouse's.
HOURSWORK	Mean number of hours worked on main job per week.
PROFESSIONAL	Dummy variable for respondent's major Census occupation.
MANAGERIAL	Dummy variable for respondent's major Census occupation.
CRAFT	Dummy variable for respondent's major Census occupation.
OPERATIVE	Dummy variable for respondent's major Census occupation.
OFFERS	"About how easy would it be for you to find a job with another employer with approximately the same income and fringe benefits you now have?" (1 = not easy at all; 3 = very easy)
UNUSED SKILLS1	A scale (Cronbach's α = .79) constructed from the following terms: "My job lets me use my skills and abilities." (4 = strongly disagree; 1 = strongly agree); "I am given a chance to do the things I do best." (4 = not at all true of job; 1 = very true)
UNUSED SKILLS2	"Do you have skills from your experience and training that you would like to be using in your work but can't use on your present job?" (1 = yes; 0 = no)
LACK RESOURCES	A scale (Cronbach's α = .72) constructed from four items of the form, "I have enough [help and equipment, information, authority, time] to get the job done" (1 = very true of job; 4 = not at all true of job).
SCREEN	"In general, how well would you say your job measures up to the sort of job you wanted when you took it?" (5 = not very much like; 1 = very much like)

[1] The *QES* records this variable as a set of class intervals; workers were assigned the midpoint of their respective classes.

ment is a distinguishing feature of the core sector (Beck, Horan, and Tolbert 1978; Hodson and Kaufman 1982), CORE is the opposite of LOSEJOB and should decrease search.

Some of the above variables may indicate the job opportunities facing a worker (e.g., CORE, AGE), but the vacancy rate as it affects the likelihood of finding a job has usually been indicated by features of the local labor market. To the extent that larger cities have a higher density of potential employers and/or more dispersed (see eq. 2) wage distributions (Black 1981), BIGCITY will increase the propensity to search. The regional dummy WEST may

capture differences in opportunities at the time of the survey, and, along with BIGCITY, adjusts current wage for cost of living differences.[6] Unemployment rate was not available,

---

[6] Although the original *QES* four-category classification of city size and region (West, North-central, Northeast, South) allowed for a finer representation of these variables, experiments showed that a single dummy for each variable sufficed to capture the systematic variation induced in search propensity. Similarly, the representation of occupational categories (Table 1) is not as fine as it might have been. The choice of which occupation dummies to include was based on the effort to

and, in any case, has had effects ranging from none (Kahn and Low 1984) to marginal (Black 1981) in other studies.

The remaining "objective" factors are self-explanatory, with two exceptions: OTHERIN-COME controls for the financial capacity to undertake search, and OTHERJOB may indicate a worker's incentive to reduce fixed work costs by consolidating jobs (Black 1981).

The remainder of Table 1 gives the measures of market and job conditions as these are known to the worker. OFFERS is a subjective estimate of the likelihood of eliciting an acceptable alternative offer in the vicinity of the critical wage—Q_1 in eq. 3. The question's qualifier—"with approximately the same income and fringe benefits"—nicely captures the concept of the critical wage.[7] Three variables—UNUSED SKILLS1, UNUSED SKILLS2, and LACK RESOURCES—are used to measure the extent to which a worker believes that the current job *fails* to fully employ his productive resources. Since these variables tap *unused* productive capacity, they should covary *directly* with the expected wage gains from search. The two UNUSED SKILLS variables are self-explanatory; LACK RESOURCES is included because the full utilization of productive capacity depends on the worker's access to means of production. Finally, SCREEN, which indicates the extent to which properties of the current job *fall short* of a worker's expectations at the time of hiring, should increase search propensity.

FINDINGS

A Baseline Search Model

I begin with a model that incorporates only the conventional indicators of a worker's job and market position. Table 2 displays the estimates of three such search functions, the only difference among them being the indicator of wage gains—LNWAGE, WAGEGAIN, and %GAIN. Judging by the *t*-ratios for these three indicators, the differences in explanatory power are minor. Nor does it matter much for the estimates of the other coefficients, with the possible exception of FEMALE and SCHOOL-ING, which wage indicator is used: the coefficients are quite stable across equations. Still, since %GAIN yields the best fit, I take MODEL III as a starting point.

Model III shows that the strongest predictors are among those variables that most directly indicate elements of the wage-gain component of the expected return to search. These include %GAIN, BENEFITS, LOSEJOB, TENURE, UNION, and AGE, all of which have the expected sign and are statistically significant. The dummy variable for supervisory status has the wrong sign, and the coefficient of FIRM-SIZE is within sampling error of zero, suggesting that the advantages of working for a large firm are adequately controlled by other variables. The coefficient of CORE indicates, as expected, that working in the core sector reduces search propensity.

Of the two measures of local labor-market conditions, BIGCITY and WEST, only the latter has an appreciable positive effect on search, presumably because of the better job opportunities in that region during the 1977 period.[8] The NONWHITE and FEMALE coefficients indicate that nonwhites and females search more than their white and male counterparts, though only the racial contrast, which is consistent with other studies (Kahn and Low 1984), is statistically significant. Of the other variables, only two are notable: OTHERJOB has the expected positive effect on search, while OPERATIVES set themselves apart from other occupational groups by their greater propensity to search.

These results closely resemble those of other studies of employed search, and match up well with empirical quit functions. This suggests that the systematic variation in search that is usually captured by objective measures is well accounted for here. In order to render the analysis more manageable, but still maintain a robust specification, only variables whose coefficients have a *t*-ratio exceeding unity are retained for the next stage.[9] Eliminating the 11 variables that do not meet this criterion yields a trivial increase in residual variation ($\Delta SSE = 1.48$, $F = 0.28$), so little is lost in explanatory power. The estimates of the trimmed equation are given in Table 3 as MODEL I, which acts as a baseline for assessing an augmented model.

An Augmented Search Model

A better representation of the mobility orientation and search decisions of workers might be

---

maximize the systematic variation accounted for by occupation while at the same time achieving parsimony.

[7] There is a minor ambiguity in the OFFERS questions insofar as "easy" may be a function not only of the density of vacancies, but of the search skill of the worker. This is partly compensated for by the inclusion of schooling and labor-force experience, which should function as controls for search skill.

[8] I do not know for a fact that opportunities actually were better in the West. But it is true that workers residing in the West are significantly more optimistic about the availability of OFFERS than workers in other regions. The consequences of this will become apparent shortly, and will render the issue of regional variation in actual opportunities moot.

[9] This practice will be used throughout the analysis to generate baseline models.

Table 2. Regression Estimates of an "Objective" Search Function ($N = 1,117$)

Independent Variables	Model I		Model II		Model III	
	β	t-ratio	β	t-ratio	β	t-ratio
Constant	2.56	9.20	2.21	10.10	2.20	10.05
Gain factors						
LNWAGE	−.103	−1.92				
WAGEGAIN			.110	1.95		
%GAIN					.598	2.23
BENEFITS	−.046	−3.67	−.047	−3.73	−.046	−3.65
LOSEJOB	.170	6.04	.169	6.01	.169	6.02
TENURE	−.034	−2.82	−.035	−2.91	−.034	−2.90
TENURE²	.001	2.30	.001	2.37	.001	2.37
AGE	−.014	−3.15	−.014	3.10	−.014	−3.07
LFEXP	.005	1.09	.004	0.85	.004	0.81
UNION	−.145	−2.79	−.166	−3.19	−.164	−3.16
SUPERVISOR	.074	1.41	.069	1.30	.069	1.31
FIRMSIZE	.023	0.76	.022	0.73	.021	0.72
CORE	−.085	−1.58	−.101	−1.90	−.099	−1.85
Market conditions						
BIGCITY	−.034	−0.62	−.050	−0.93	−.049	−0.91
WEST	.112	1.89	.102	1.72	.103	1.75
Other factors						
NONWHITE	.145	1.92	.158	2.11	.156	2.08
FEMALE	.033	0.53	.069	1.20	.068	1.19
OTHERINCOME	−.004	−0.14	−.002	−0.08	−.003	−0.12
OTHERJOB	.183	2.95	.183	2.95	.184	2.97
HOURSWORK	.005	0.20	.005	0.20	.005	0.22
SCHOOLING	.012	1.09	.006	0.56	.006	0.55
MARRIAGE	−.006	−0.10	−.016	−0.26	−.015	−0.25
KIDS < 6	−.056	−1.38	−.058	−1.43	−.057	−1.41
KIDS < 18	−.008	−0.38	−.011	0.53	−.012	−0.56
PROFESSIONAL	−.008	−0.11	.023	0.31	−.024	−0.23
MANAGERIAL	−.004	−0.05	.033	0.38	−.032	−0.37
CRAFT	.052	0.72	.036	0.49	.037	0.51
OPERATIVE	.138	2.08	.136	2.05	.134	2.01
$(N-k-1)\hat{\sigma}^2$		473.6		473.5		472.9
d.f.		990		990		990
R^2		195		.195		.196
F		9.21		9.21		9.27

achieved by directly measuring workers' informed assessments of their job opportunities and potential wage gains. In place of aggregate indicators of local labor demand—in this case WEST and BIGCITY—I suggested measuring workers' subjective estimates of the likelihood of eliciting an acceptable offer, here tapped by OFFERS. As an alternative to the usual imputed money measures of wage gain—in this case %GAIN—I suggested that workers may infer their potential wage gains by assessing the extent to which the current job fails to fully employ their skills and abilities. This yielded the variables UNUSED SKILLS1, UNUSED SKILLS2, and LACK RESOURCES.

The results of augmenting the conventional model with these variables is given by Model II of Table 3. Introducing these variables yields a sizable reduction in residual variation ($\Delta SSE = -40$, $F = 22.5$), with all the coefficients in the expected direction and significant. Judging by the t-ratios, these variables are also among the strongest predictors in the model. Workers who believe it would be easy to find a job offering rewards comparable to those of their current job, and who believe that their current job does not fully employ their productive resources, are much more likely to search than other workers. Furthermore, controlling these variables has resulted in drastic reductions in the coefficients of their "objective" counterparts: the coefficients of %GAIN and WEST are reduced by 75 percent and 50 percent, respectively, and are now less than their standard errors.

These results are promising, but the model of equation 2 can be pushed even further. Consider the partial derivatives of R with respect to the vacancies terms $q(v)$ and the wage gap factor $(W_c - \mu_a)$:

$$\partial R/\partial q(v) = \sigma^2 \dot{f}(W_c) - (W_c - \mu_a)\phi(W_c > 0 \quad (4)$$

$$\partial R/\partial(W_c - \mu_a) = -q(v)\phi(W_c) \leq 0. \quad (5)$$

Table 3. Regression Estimates of an "Objective" Search Function Augmented by Information-Based Indicators of Opportunity and Gain ($N = 1,024$)

Independent Variables	Model I β	Model I t-ratio	Model II β	Model II t-ratio	Model III β	Model III t-ratio
Constant	2.17	16.11	2.62	13.50	2.05	7.29
Objective measures						
%GAIN	.489	1.91	.166	0.67	.150	0.61
BENEFITS	−.044	−3.78	−.042	−3.72	−.041	−3.63
LOSEJOB	.167	6.04	.146	5.39	.152	5.60
TENURE	−.036	−3.11	−.029	−2.53	−.027	−2.44
TENURE²	.001	2.56	.001	2.19	.001	2.14
AGE	−.011	−4.91	−.009	4.13	−.009	−4.19
UNION	−.160	−3.18	−.154	−3.12	−.157	−3.20
SUPERVISOR	.046	0.94	.066	1.39	.067	1.42
CORE	−.094	−1.83	−.071	−1.42	−.066	−1.33
WEST	.104	1.80	.053	0.96	.048	0.86
NONWHITE	.154	2.12	.156	2.24	.151	2.17
FEMALE	.054	1.13	.082	1.76	.077	1.66
OTHERJOB	.185	3.02	.150	2.55	.149	2.54
KIDS < 6	−.058	−1.51	−.045	−1.22	−.044	−1.20
OPERATIVE	.119	2.01	.093	1.60	.089	1.54
Information measures						
OFFERS			.077	5.23	.308	3.60
UNUSED SKILLS1			.044	3.33	.016	0.70
UNUSED SKILLS2			.180	3.99	.080	0.92
LACK RESOURCES			.038	3.90	.007	0.41
OFFERS						
×UNUSED SKILLS1					.013	1.69
×UNUSED SKILLS2					.040	1.38
×LACK RESOURCES					.012	2.02
$(N-k-1)\hat{\sigma}^2$		478.7		439.3		433.3
d.f.		1008		1004		1001
R^2		.191		.258		.268
F		15.9		18.36		16.65

Eq. 4 shows that the effect of $q(v)$ on search is only partly contingent on the gap between the critical wage and the mean market-wage offer: even if $(W_c - \mu_a) = 0$, $\partial R/\partial q(v) > 0$. In other words, the effect of vacancies is only partly due to its interaction with the wage-gap term. Eq. 5 shows that the reverse does not hold: the effect of $(W_c - \mu_a)$ is wholly contingent on its interaction with vacancies, for if $q(v) = 0$, then $\partial R/\partial(W_c - \mu_a) = 0$.

These ideas may be represented in the relevant part of the empirical model by

$$R = \beta \text{OFFERS} + \sum_{j=1}^{3} \alpha_j[\text{OFFERS} \times X_j] \quad (6)$$

where the X_j are the relative productivity variables (UNUSED SKILLS1, etc.) and all parameters are positive. In this model, the relativity productivity indicators take the place of $(W_c - \mu_a)$, and OFFERS, a measure of $Q_1 = q(v)f(W_c)$, is used to approximate $q(v)$. This specification, an alternative to Model II of Table 3, expresses two plausible ideas. First, workers who believe their productive resources are not fully employed on the current job, and that,

therefore, wage gains may be achieved by changing jobs, will search *only if* they also believe that the likelihood of eliciting an acceptable offer is reasonably good. In other words, potential productivity/wage gains affect search only when accompanied by perceived job opportunities. Second, the likelihood of eliciting acceptable offers will instigate search even among workers who believe their skills are already fully employed ($\beta > 0$ even when $X_j = 0$). Naturally, the effect of acceptable offers will be even greater (i.e., $\beta + \alpha_j$) among workers whose skills are not fully employed.

Augmenting the additive Model II with the interaction terms yields Model III of Table 3. The results rule in favor of the multiplicative specification. The set of three interaction terms significantly reduce the residual variation ($\Delta SSE = -5.95$, $F = 4.58$, $p < .004$), and the individual coefficients are all in the right direction and have t-ratios that are significant with directional tests and allowance for multicollinearity.[10] Furthermore, the coefficient of the

_____

[10] The next model shows that when the additive terms for relative productivity are dropped from the equation, all the coefficients of the interaction terms have t-ratios

Table 4. Regression of SCREEN on Selected Independent Variables from the Search Equation ($N = 1,142$)

Independent Variables	β	t-ratio
Constant	4.56	16.41
STARTEXP[a]	−.001	−0.38
CORE	.111	1.28
UNION	−.174	−2.05
NONWHITE	.306	2.46
FEMALE	.020	0.25
OPERATIVES	.309	2.99
UNUSED SKILLS1	.178	7.68
UNUSED SKILLS2	.452	5.58
LACK RESOURCES	.085	4.83
R^2		.168
F		25.4

[a] This variable is an estimate of the worker's labor-force experience at the time he or she was hired by his current employer.

additive term for OFFERS is significant, but the coefficients of the additive terms in relative productivity are, both as a group ($\Delta SSE = 1.02$, $F = 0.79$, $p = .50$) and taken singly, well within sampling error of zero.

Screen Quality and Search

This section examines how the quality of the initial screen affects the propensity to search. A related issue is whether systematic differences in search propensity across types of workers and market conditions partly reflect differences in screen quality and hence in the risk of mismatch. We have seen, for example, that white workers and male workers are less likely to search than their black or female counterparts. Is this partly attributable to informational or signalling differences that reduce the risk of mismatch for males and whites? The same kind of question applies to the finding that core workers are slightly less likely and union workers much less likely to search. Finally, there is the question whether the worker's relative productivity is itself an experience property subject to an imperfect screen.

Table 4 shows the results of regressing SCREEN on selected independent variables from the search function. This regression is meant only as a concise representation of the net associations relevant for the analysis of search. The estimates show that whites and union workers are slightly less likely to say the job is not as good as expected, the speculation being that these workers have better information about the current job prior to hiring. On the other hand, "operatives" are more likely than workers

exceeding four. This change is due mainly to a decrease in the standard errors rather than an increase in the magnitude of the coefficients.

in other occupations to report that their job falls short of original expectations. And as these estimates clearly indicate, the quality of the initial screen is significantly implicated in worker assessments of their relative productivity. Workers who feel that their productive resources are not fully employed on the current job are much more likely to say the job does not measure up to expectations. For some workers, then, their relative productivity on a job appears to be learned through working, rather than immediately inferred when the match is formed.

MODEL II in Table 5 displays the results of including an additive term for SCREEN and a term for its interaction with TENURE in the search function (MODEL I serves as a baseline). This specification was based on the expectation, borne out by the results, that the initial screen is most relevant to the search decision early in the match and becomes increasingly irrelevant as tenure increases. By the same reasoning, after a certain number of years the quality of the initial screen might not matter at all. This suggests a discontinuous regression relation in which the declining positive effect of SCREEN is limited to a lower range of tenure, beyond which it is forced to zero. I therefore fitted a series of models that imposed a discontinuity at various points along the range of tenure. The fit that minimized residual variation, displayed as MODEL III, constrains the declining positive effect of SCREEN to workers with length of service less than seven years. This brings a slight but costless reduction in residual variation over Model II.

The other coefficients of MODEL III indicate that the quality of the screen is not seriously implicated in the estimates of the effects of related variables. Comparing MODEL III to MODEL I shows that controlling for SCREEN does, as expected, reduce the coefficients of the relative productivity variables, but the reductions are small and the coefficients remain highly significant.[11] Similarly, the coefficients of the other correlates of SCREEN (e.g., UNION and NONWHITE) identified above undergo only minor changes, so that the effects of these factors do not appear to be attributable to informational differences at the time of hiring. The one exception occurs with respect to the coefficient of OPERATIVES, which is reduced by almost 40 percent with controls for SCREEN. It appears that the slight tendency for these workers to search more than those in other occupations may be due to their greater

_____

[11] The coefficient of OFFERS also declines, but this is due to its correlation with the interaction terms. The zero-order correlation between OFFERS and SCREEN is itself indistinguishable from zero.

Table 5. Regression Estimates of a Search Function Augmented by Controls for Screen Quality ($N = 1,097$)

Independent Variables	Model I		Model II		Model III	
	β	t-ratio	β	t-ratio	β	t-ratio
Constant	1.92	14.45	1.53	11.08	1.47	10.49
BENEFITS	−.041	−3.92	−.040	−3.92	−.039	−3.90
LOSEJOB	.157	6.17	.151	6.08	.149	6.02
TENURE1	−.028	−2.60	−.003	−0.28	.022	1.58
TENURE²	.001	2.17	.001	1.97	−.000	−0.69
AGE	−.009	−4.70	−.009	4.49	−.008	−4.46
UNION	−.141	−3.00	−.133	−2.90	−.134	−2.91
SUPERVISOR	.060	1.32	.072	1.63	.076	1.73
CORE	−.055	−1.17	−.070	−1.50	−.067	−1.44
NONWHITE	.176	2.67	.144	2.25	.146	2.27
FEMALE	.075	1.69	.084	1.94	.089	2.05
OTHERJOB	.145	2.54	.143	2.56	.139	2.50
KIDS < 6	−.046	−1.30	−.046	−1.34	−.044	−1.31
OPERATIVE	.071	1.29	.040	0.75	.043	0.80
OFFERS	.351	7.55	.293	6.32	.286	6.23
×UNUSED SKILLS1	.018	4.32	.013	3.12	.013	3.01
×UNUSED SKILLS2	.065	4.53	.054	3.79	.053	3.75
×LACK RESOURCES	.014	4.40	.012	3.86	.011	3.78
Screen Variables						
SCREEN			.160	7.75		
×TENURE			−.011	−5.12		
DSCREEN					.186	7.77
×TENURE					−.019	−4.51
$(N-k-1)\hat{\sigma}^2$		464.3		439.7		438.1
d.f.		1079		1077		1077
R^2		.265		.304		.307
F		22.9		24.8		25.1

difficulty in accurately anticipating properties of prospective jobs.

Sex Differences in Search Propensity

MODEL I of Table 6 serves as a baseline for examining sex differences in search propensity. The point estimate (.079) of the net sex difference in search propensity is close to the gross difference in means (.085), so that the model has not succeeded in explaining why women are more prone to embark on search than men. To determine whether this was all that needed to be explained, I carried out a test of sex differences in coefficient vectors (conditional on an intercept difference). Dropping the equality constraint of MODEL I reduced the residual variation from 440 to 434 on 16 degrees of freedom, yielding an F-ratio of 0.82. Furthermore, only 3 of the 16 pairwise coefficient contrasts yielded t-ratios greater than 1 (all ≤ 1.56). Excepting the difference in intercepts, then, it looks as if a single search function governs the behavior of both sexes.

It is already clear that the additive sex difference is not due to systematic and pervasive disparities in acceptable offers, productivity/ wage gains, or screen quality, since these are in the model. Controlling these factors does not eliminate the sex effect because male-female

differences on these variables are either nonexistent or countervailing to begin with. Hence, there is no discernible sex difference in either OFFERS ($\bar{X}_m - \bar{X}_f = .026$, $t = 0.46$) or SCREEN ($\bar{X}_m - \bar{X}_f = .039$, $t = 0.45$). Female mean levels of UNUSED SKILLS are significantly higher than those of men ($t \sim 2$), but their mean level of LACK RESOURCES is significantly lower ($t = 2.52$).

An alternative explanation of the sex effect is needed. One possibility is that it expresses a sex-linked and perhaps unpredictable job condition that affects a small proportion of women, promotes search, and is powerful enough to pull the overall female propensity to search above that of males. As suggested earlier, the experience of sexual discrimination may be such a job condition, and certain facts underscore this conjecture. First, discrimination is reported by only 13 percent of the women in the sample.[12] Second, women reporting discrimination evince a much strong propensity to search ($\bar{X} = 1.83$) than other females ($\bar{X} = 1.51$) and males ($\bar{X} = 1.46$), with the latter difference nonsignificant. Third, the experience of discrimination is fairly unpredictable, at least in the sense that women

---

[12] The discrimination question is, "Do you feel in any way discriminated against on your job because you are a woman?" (1 = yes, 0 = no).

Table 6. Regression Estimates for the Effect of Perceived Sexual Discrimination on the Search Prosperity of Female Workers ($N = 1,097$)

Independent Variables	Model I		Model II	
	β	t-ratio	β	t-ratio
Constant	1.51	11.19	1.44	10.56
BENEFITS	−.041	−4.03	−.042	−4.15
LOSEJOB	.150	6.07	.148	5.97
TENURE	.013	2.93	.013	2.85
AGE	−.009	−4.55	−.008	−4.40
UNION	−.132	−2.90	−.130	−2.88
SUPERVISOR	.068	1.58	.074	1.71
CORE	−.071	−1.55	−.072	−1.56
NONWHITE	.145	2.26	.148	2.31
OTHERJOB	.141	2.55	.139	2.51
KIDS < 6	−.046	−1.35	−.044	−1.29
OFFERS	.293	6.35	.291	6.33
×UNUSED SKILLS1	.014	3.29	.014	3.41
×UNUSED SKILLS2	.053	3.77	.053	3.75
×LACK RESOURCES	.011	3.78	.011	3.64
DSCREEN	.165	8.10	.163	8.00
×TENURE	−.012	−6.12	−.012	−6.15
FEMALE	.079	1.85		
BIAS			.270	2.93
NO BIAS			.048	1.08
$(N-k-1)\hat{\sigma}^2$		440.1		437.6
d.f.		1079		1078
R^2		.304		.307
F		27.7		26.6

reporting discrimination show significantly poorer levels of SCREEN quality ($\bar{X} = 2.72$) than either other women ($\bar{X} = 2.19$) or men ($\bar{X} = 2.22$).

The effect of discrimination may be examined by forming two female dummies, one for females reporting discrimination (FEMALE BIAS) and one for other females (FEMALE NO BIAS). The estimates of MODEL II show that sexual discrimination goes a long way toward explaining the sex difference in search. The net difference between women experiencing discrimination and males is .270 ($p < .004$), which compares to the gross difference cited above of .370. The decline is due partly to controls for SCREEN, but mainly to the fact that women who report discrimination are significantly more pessimistic ($\bar{X} = 2.24$) than either other females ($\bar{X} = 2.52$) or males ($\bar{X} = 2.51$) about the chances of eliciting reward-equivalent OFFERS from other employers. Furthermore, women who experience discrimination show a significantly higher propensity to search than even other women (.270 − .048 = .222, $p < .01$), while the difference between other women and men is nonsignificant. This evidence argues against a *general* sex difference in search propensity and for a much narrower conclusion: the experience of discrimination has a powerful *direct* effect on the propensity of women workers to search. This finding augments Osterman's (1982) work linking discrimination to female quit rates.

Specification Analysis

Two final issues concern the statistical specification of the basic model used throughout the analysis. The first has to do with the endogeneity of market information. I have assumed that market information about the availability of acceptable offers is predetermined with respect to the search decision, while realizing that this information will be adaptively updated once search is under way. The idea is that workers passively receive from a variety of sources—friends, coworkers, mass media—market information, which they turn into subjective estimates of job opportunities. These estimates are then used to compute the expected returns to search, which in turn determines the propensity to search.

While this portrait may describe most workers, an alternative causal sequence may characterize the decision process in some nontrivial proportion of cases. Some workers may actively search before converting market information into estimates of available offers. Having already invested in search, these workers may then become overly optimistic about locating an acceptable offer, yielding the positive association between search and perceived offers. Of course, this need not be a one-way process: a growing investment in search and a growing optimism about opportunities may be mutually reinforcing. In either case, the terms involving OFFERS in the various equations might be

Table 7. Alternative Estimates of the Final Search Model

Independent Variables	OLS		2-SLS		Ordered Logistic	
	β	t-ratio	β	t-ratio	β	t-ratio
Constant	1.44	10.41	1.27	3.77	1.23	9.8
BENEFITS	−.041	−4.03	−.036	−2.65	−.125	−4.18
LOSEJOB	.150	5.98	.149	5.09	.545	7.32
TENURE	.013	2.96	.013	2.84	.022	1.37
AGE	−.009	−4.49	−.008	−4.15	−.031	−5.25
UNION	−.126,	−2.76	−.102	−1.68	−.564	−3.79
SUPERVISOR	.081	1.85	.086	1.78	.252	1.84
CORE	−.073	−1.58	−.061	−1.21	−.256	−1.88
NONWHITE	.154	2.37	.156	2.38	.555	3.03
OTHERJOB	.144	2.59	.125	1.92	.515	3.24
KIDS < 6	−.043	−1.25	−.036	−1.02	−.160	−1.57
OFFERS	.288	6.20	.406	3.36	.922	7.06
× UNUSED SKILLS1	.014	3.40	.023	3.35	.041	3.52
× UNUSED SKILLS2	.051	3.56	.058	3.42	.155	3.92
× LACK RESOURCES	.010	3.44	.011	3.08	.035	4.07
DSCREEN	.167	8.10	.157	7.12	.474	8.07
× TENURE	−.012	−6.25	−.012	−5.74	−.032	−4.73
FEMALE BIAS	.254	2.86	.269	2.87	.868	3.43
FEMALE NO BIAS	.031	0.68	.035	0.78	.129	0.92
$(N-k-1)\hat{\sigma}^2$	429.8		436.9		1781[a]	
N	1077		1077		2196[b]	
R^2	.310		.294		.268[c]	
F	27.7		26.6		36.2[d]	

[a] This figure is $-2 \times$ log likelihood.

[b] This figure is so large because the procedure used to estimate the model doubles the number (N = 1,098) of observations. See note 14.

[c] This is the so-called pseudo-R^2, computed as $(L_1 - L_2)/L$, where L_1 and L_2 are $-2 \times$ log likelihood for the null and fitted models, respectively.

[d] This is the likelihood ratio chi-square statistic for the test of the null hypothesis that all coefficients are zero.

positively related to the disturbances, rendering the coefficients upwardly biased.

To correct for such bias, the search equation was estimated by two-stage least squares.[13] Table 7 presents these estimates and the corresponding ordinary least-squares estimates. Comparing the estimates, the 2-SLS coefficients are the same or lower than their OLS counterparts in most cases, with an important exception: the estimates of coefficients of terms involving OFFERS are larger, in some cases much larger, than the OLS estimates. These results boost the credibility of the original specification and leave all major findings firmly intact.

Finally, there is the issue of modelling the *ordinal* measure of search propensity by linear regression. An alternative would have been to use a model that treats the ordinal responses as discrete categorical expressions of an underlying continuous distribution of search propensities. One such model is the ordered logistic (Cox 1970; McCullagh 1980), the estimates of which are also given in Table 7.[14] Although the coefficients are not comparable to those of the linear regressions, the pattern and magnitudes of the t-ratios are very similar across the models. Once again, the findings are largely impervious to changes in the statistical model.

CONCLUSION

Few would contest the idea that the decision to change jobs is guided by the information available to workers about their job and market position. Yet such a conceptualization has not

[13] The "reduced form" equation that yielded the predicted values of OFFERS to be used in the second stage included all the exogenous variables in the search equation (Model II, Table 6), plus five occupation dummies, firm size, and the region dummy WEST. Technically speaking, then, the structural equation for search propensity is overidentified. In the second stage, the observed OFFERS was replaced by predicted OFFERS in both the additive and interaction terms.

[14] The ordered logistic was estimated by a dichotomous logit procedure described by Winship and Mare (1984). They claim that this procedure yields estimates that are within 10 percent of those obtained by maximum likelihood, which squares with my experience. The procedure means doubling the number of observations (from 1,098 to 2,196), which imparts a slight, usually trivial, downward bias to the standard errors.

been an integral feature of current structuralist models of job mobility. Nevertheless, structuralist models are quite compatible with and actually invite an approach that attends more closely to the action orientation of workers. Nothing said here suggests otherwise. On the contrary, the preceding analysis is built around a model that dimensionalizes workers' subjective assessments of their job/market position and the search-decision problem in terms of the principal structural components of the job-mobility models that Sørensen's work has spawned.

The rationale underlying the analysis is simple. If exogenously generated opportunities and structurally constrained gains are as significant as current models assume, then these structural forces ought to be evidenced in the subjective orientation workers bring to mobility in general and the search decision in particular. The findings reported here provide such evidence, buttressing structuralist assumptions. Furthermore, the findings tie opportunities and gains to the search decision in a way that adds a new twist and refinement to thinking about the process generating an orientation to mobility. The twist involves conceptually linking worker estimates of potential gains to their subjective assessments of the extent to which the current job fully employs their skills and abilities. The empirical implications of this conceptualization are fully supported by the findings. The refinement concerns the asymmetrical multiplicative specification of how subjective assessments of market opportunities and potential achievement gains combine to shape workers' orientation to job change. On this count, too, the findings are supportive, arguing against a simple additive specification.

While this much of the analysis remains firmly rooted in structuralist formulations of the job-mobility process, the treatment of screen quality and discrimination-based sex differences constitute an extension. The quality of the initial screen conditions a worker's risk of mismatch once in the job and, as such, has constituted a potential explanation of observed disparities in mobility rates across types of workers and firm/market conditions. I have found that while screen quality does indeed bear on workers' incentive to search, it is not profoundly implicated in the disparities in search propensity observed across types of workers or firm/market conditions. Nor is screen quality, which figures in recent economic explanations of gender disparities in turnover, implicated in any general way in the finding that women express a stronger propensity to search than men. Rather, the sex difference in search propensity was largely explained by a sex-linked condition of work: sexual discrimination is a powerful, direct stimulus to search for the small proportion of women who report experiencing it on the job. It is noteworthy that, were it not for sexual discrimination, the structures of the functions explaining the search orientations of male and female workers would be statistically indistinguishable.

These findings, and the framework surrounding them, point the study of mobility processes toward the action- and choice-based approach recently advocated by Coleman (1986). Yet it would be a mistake to see this as an invitation to · abandon structuralist formulations, since such formulations are at the heart of the present work. Moreover, I strongly propose that further efforts in the direction taken here must devote more attention to mapping systematically the way institutional, organizational, and market conditions mix to shape both worker judgments about opportunities and wage prospects and the quality of the initial screen. Little is known about the process by which workers arrive at subjective assessments on which their mobility orientations are based, nor about how such assessments are tied to the broader structural conditions that constrain action. Although most of the variables that usually appear as indicators of "structure" in job shift studies have been included here, the task remains to link these to the subjective assessments that inform mobility activity.

But the obligation runs both ways. Structuralist formulations need to do more to link structure to the thing it constrains—the decision calculus governing choice. Closer attention to the intervening action mechanisms that link structure to job change will surely advance theory that integrates knowledge of how macroscopic social structures condition mobility processes. As Carroll and Mayer have observed, research to this point suggests the complementarity of various structuralist perspectives, and "what is needed now is integrative theory" (1986, p. 336). I concur, but part company when they voice the widely shared view that integration may be achieved by deploying still more structural concepts. This strategy bypasses action and delays the construction of concepts that will bridge the gap between structural phenomena and mobility decisions. As Coleman (1987) and Sewell (1987) have persuasively argued, explicating such macro-micro relations is a critical but underdeveloped dimension of modern social research. A second problem with the Carroll-Mayer strategy is that it does not fairly credit the power of existing structural concepts. In my opinion, the concepts showing the most promise for integrating structuralist perspectives have already been deployed— opportunities and achievement gains. The special appeal of these concepts is that they are directly applicable to individual choice and, at

that level, provide a fruitful basis for modelling the pursuit of mobility by knowing and acting workers.

REFERENCES

Barnes, William F. and Ethel Jones. 1974. "Differences in Male and Female Quitting." *Journal of Human Resources* 9:439–51.

Baron, James N. and William T. Bielby. 1984. "The Organization of Work in a Segmented Economy." *American Sociological Review* 49:454–73.

Barron, John M. and Stephen McCafferty. 1977. "Job Search, Labor Supply, and the Quit Decision: Theory and Evidence." *American Economic Review* 67:683–91.

Beck, E.M., Patrick M. Horan, and Charles M. Tolbert. 1978. "Stratification in a Dual Economy: A Sectoral Model of Earnings Determination." *American Sociological Review* 43:704–20.

Black, Matthew. 1980. "Pecuniary Implications of On-the-Job Search and Quit Activity." *Review of Economics and Statistics* 62:222–29.

———. 1981. "An Empirical Test of the Theory of On-the-Job Search." *Journal of Human Resources* 16:129–40.

Blau, Francine D. and Lawrence M. Kahn. 1981. "Race and Sex Differences in Quits by Young Workers." *Industrial and Labor Relations Review* 34:563–77.

Bridges, William P. and Wayne J. Villemez. 1986. "Informal Hiring and Income in the Labor Market." *American Sociological Review* 51:574–82.

Burdett, Kenneth. 1978. "A Theory of Employee Search and Quit Rates." *American Economic Review* 68:212–20.

Carroll, Glenn R. and Karl Ulrich Mayer. 1986. "Job-Shift Patterns in the Federal Republic of Germany: The Effects of Social Class, Industrial Sector, and Organization Size." *American Sociological Review* 51:323–41.

Coleman, James S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91:1309–35.

———. 1987. "Actors and Action in Social History and Social Theory: Reply to Sewell." *American Journal of Sociology* 93:172–75.

Cox, D.R. 1970. *The Analysis of Binary Data*. London: Chapman and Hall.

Datcher, Linda. 1983. "The Impact of Informal Networks on Quit Behavior." *Review of Economics and Statistics* 65:491–95.

DiPrete, Thomas A. 1981. "Unemployment Over the Life Cycle: Racial Differences and the Effect of Changing Economic Conditions." *American Journal of Sociology* 87:286–307.

Farber, Henry S. and Daniel H. Saks. 1980. "Why Workers Want Unions: The Role of Relative Wages and Job Characteristics." *Journal of Political Economy* 88:349–69.

Felmlee, Diane H. 1982. "Women's Job Mobility Processes Within and Between Employers." *American Sociological Review* 47:142–50.

Freeman, Richard B. 1980. "The Exit-voice Tradeoff in the Labor Market." *Quarterly Journal of Economics* 94:643–73.

Garen, John E. 1985. "Worker Heterogeneity, Job Screening, and Firm Size." *Journal of Political Economy* 93:715–39.

Granovetter, Mark. 1974. *Getting a Job: A Study of Contacts and Careers*. Cambridge, MA: Harvard University Press.

Gronau, Reuben. 1971. "Information and Frictional Unemployment." *American Economic Review* 61:290–301.

Hodson, Randy and Robert L. Kaufman. 1982. "Economic Dualism: A Critical Review." *American Sociological Review* 47:727–39.

Hutchens, Robert. 1986. "Delayed Payment Contracts and a Firm's Propensity to Hire Older Workers." *Journal of Labor Economics* 4:439–57.

Jovanovic, Boyan. 1979a. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87:972–90.

———. 1979b. "Firm-specific Capital and Turnover." *Journal of Political Economy* 87:1246–60.

———. 1984. "Matching, Turnover, and Unemployment," *Journal of Political Economy* 92:108–22.

Kahn, Lawrence M. and Stuart A. Low. 1982. "The Relative Effects of Employed and Unemployed Job Search." *Review of Economics and Statistics* 64:234–41.

———. 1984. "An Empirical Model of Employed Search, Unemployed Search, and Nonsearch." *Journal of Human Resources* 19:104–17.

Lippman, Steven A. and John J. McCall. 1976. "The Economics of Job Search: A Survey." *Economic Inquiry* 14:155–89 and 347–68.

McCall, J.J. 1981. "The Economics of Belated Information." *International Economic Review* 22:135–46.

———. 1970. "Economics of Information and Job Search." *Quarterly Journal of Economics* 84:113–26.

McCullagh, P. 1980. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society, B*, 42:109–42.

Meitzen, Mark E. 1986. "Differences in Male and Female Job-quitting Behavior." *Journal of Labor Economics* 4:151–67.

Osterman, Paul. 1982. "Affirmative Action and Opportunity: A Study of Female Quit Rates." *Review of Economics and Statistics* 64:604–12.

Parsons, Donald O. 1973. "Quit Rates Over Time: A Search and Information Approach." *American Economic Review* 63:390–401.

———. 1977. "Models of Labor Market Turnover: A Theoretical and Empirical Survey." Pp. 185–223 in *Research in Labor Economics*, vol. 1, edited by Ronald Ehrenberg, Greenwich, CT: JAI Press.

Quinn, Robert P. and Graham Stanley. 1979. *Quality of Employment Survey, 1977: Cross-section* [mrdf]. ICPSR ed. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor].

Sewell, William H., Jr. 1987. "Theory of Action, Dialectic, and History: Comment on Coleman." *American Journal of Sociology* 93:166–72.

Sørensen, Aage B. 1977. "The Structure of Inequality and the Process of Attainment." *American Sociological Review* 42:965–78.

———. 1979. "Intergenerational Status Attainment: A Model and a Metric." *American Journal of Sociology* 85:361–84.

Sørensen, Aage B. and Nancy Brandon Tuma. 1981. "Labor Market Structures and Job Mobility." In *Research in Social Stratification and Mobility*, vol. 1, edited by Robert Robinson. Greenwich, CT: JAI Press.

Stigler, George. J. 1962. "Information in the Labor Market." *Journal of Political Economy* 69:312–25.

Stoikov, Vladimir and Robert Raimon. 1968. "Determinants of the Differences in the Quit Rate Among Industries." *American Economic Review* 58:1283–98.

Tuma, Nancy Brandon. 1976. "Rewards, Resources, and the Rate of Mobility: A Nonstationary Multivariate Stochastic Model." *American Sociological Review* 41:338–60.

Viscusi, W. Kip. 1980. "Sex Differences in Worker Quitting." *Review of Economics and Statistics* 62:388–98.

Viscusi, W. Kip and Charles J. O'Connor. 1984.

"Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?" *American Economic Review* 74:942–56.

Waite, Linda J. and Sue E. Berryman. 1986. "Job Stability Among Young Women: A Comparison of Traditional and Nontraditional Occupations." *American Journal of Sociology* 92:568–95.

Winship, Christopher and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49:512–25.

GENDER AND PROMOTION IN SEGMENTED JOB LADDER SYSTEMS*

THOMAS A. DIPRETE
University of Chicago

WHITMAN T. SOULE
Northwestern University

White-collar internal labor markets are typically segmented in various ways. Job ladders, job ladder groupings, and tiers create a complicated set of boundaries that make advancement contingent on such factors as the chances for advancement in a job ladder, the chances to move to other job ladders that offer greater opportunity, and the distribution of preferred statuses that influence promotion outcomes. This paper shows how these contingencies created gender differences in promotion rates in the federal civil service during the middle 1970s. Gender promotions varied by level. The greatest female disadvantage occurred near the boundary between the lower- and upper-tier grades. No gender difference was found in the higher grades, and no zero-order difference existed in the lower grades. Gender differences in grade promotion were largely explained by differences in ladder-specific opportunities. But significant gender differences in promotion from the lower- to the upper-tier remained, even when personal attributes and organizational location were controlled.*

INTRODUCTION

During the past half-century, both business and government have developed increasingly formal personnel policies that limit managerial discretion in assignment and promotion decisions. By limiting discretion, these policies presumably reduced overt employment discrimination against women within these organizations. The affirmative-action initiatives of the late 1960s and 1970s may have reduced discriminatory practices still further (O'Farrell and Harlan 1984; Burstein 1985; DiPrete 1987). Despite these developments, gender segregation and gender inequality persist in large firms as well as small. This segregation is one aspect of a tiered or segmented system of job ladders that characterizes the organization of white-collar work in bureaucratized organizations (Piore 1976; Althauser and Kalleberg 1981).

Research on the extent of sex segregation (e.g., Beller 1982; Roos 1985; Bielby and Baron 1984) and on the association between sex segregation and career advancement (e.g., Wolf and Fligstein 1979; Rosenfeld 1983) has begun to clarify the links between labor-market structure and gender inequality. Sex segregation disadvantages women. It limits their initial placement to job ladders that offer relatively little opportunity for further advancement, and it places certain jobs off-limits to women in contrast with equally qualified men.

Despite this progress, however, our models still need substantial refinement. Scholarship about the effects of internal labor markets on advancement is often based on indirect evidence, or on untested and vaguely specified theories of how internal labor markets work. The extent to which gender differences in promotion depend on initial placement, or on unequal opportunity for job-ladder shifts, are largely unanswered questions. The amount of variation in these attributes with organizational and institutional variables is only beginning to be explored (Baron, Davis-Black, and Bielby 1986). A principal reason for the current condition is the lack of the necessary firm-level data that combines detailed information about job placement with information on personnel flows. Such data is needed not only to test general conceptions about how job ladders work, but also to improve and refine theorizing about them (Althauser 1987).

In this paper, we present the results of a case study of promotion rates of women in the federal government. We argue that advancement within a segmented system of job ladders depends on a number of contingencies, any of which could produce gender inequality in outcomes. We show that these contingencies create gender differences in promotion rates that vary by level in a systematic way. While the data for this paper are drawn only from the federal civil service, the similarities between the

white-collar internal labor markets of government agencies and other large bureaucratized firms suggest that the results have wider applicability.

POTENTIAL SOURCES OF GENDER DIFFERENCES IN PROMOTION RATES WITHIN ORGANIZATIONS

The sex-segregation approach to gender differences in career advancement assumes that unequal initial placement, combined with mobility barriers to subsequent advancement, cause gender inequality. To test this approach, we embed these potential sources of inequality in a broader model of the promotion process. Gender differences in advancement can be attributed to a number of factors: (1) gender differences in levels of, or rates of return to, human capital; (2) gender differences in distribution across job ladders, what is usually referred to as sex segregation; (3) gender differences in preferred organizational statuses other than rank and the job ladder of one's job; (4) unequal rates of mobility from job ladders with limited opportunity to those with greater opportunity; and (5) gender differences that persist when human capital and organizational location are controlled. We elaborate on specification issues in this section of the paper.

The probability of advancing up a job ladder depends on the level of opportunity available for such advancement. Measures such as the percent female or whether a job is in the upper or lower tier are associated with the room for advancement on a given job ladder, but neither measure is perfect. A better measure would be based on the ratio of vacancies to the number of competitors. While such detailed information is almost never available, a good substitute is the multiple grade ratio (MGR) (Stewman and Konda 1983), which is based on the ratio of workers at one level of a hierarchy to those at higher levels. The MGR has usually been applied to organizations as a whole, but it can be modified to apply to individual job ladders.

Aside from rank or location on a job ladder, an employee's advancement rate can be affected by the character of the employee's relationship with the organization. For example, part-time employees would not be expected to advance as rapidly as full-time ones, perhaps because the former acquire on-the-job training at a lower rate, or perhaps because their progress is hampered by formal personnel procedures. Other types of "irregular" employment relationships exist. Civil-service systems frequently vest a large portion of white-collar employees with a form of tenure, while also employing "irregular" employees, i.e., those not tenured or on a

"tenure-track."[1] These irregulars may work part-time, or they may have a temporary or otherwise "untenured," though indefinite, full-time job. While grade advancement for irregular employees is not impossible, it is probably slower than for regular employees.

While tenure status and job ladder generally have powerful effects on advancement, obstacles they may present can be overcome if the organization allows employees to change these statuses. The frequency of job-ladder shifts probably varies by organization, as do other characteristics of job ladders (e.g., Baron et. al. 1986). For employees in organizations with partially permeable ladders, the probability of a shift arguably depends on two factors: an employee's opportunity for further advancement on the current job ladder, and transferability of skills from one ladder to another.

Low job-ladder ceilings are a principal reason why employees would want to shift ladders. Consequently, it is reasonable to hypothesize a negative relationship between opportunity on one's job ladder and the probability of job-ladder shifts. But this relationship describes only employee preferences. The ability of employees to realize their preferences depends on management's evaluation of their qualifications for other jobs. This judgment is, in part, a function of the personal characteristics of the employee. But it also depends on the jobs in question.

In some types of work, job ladder boundaries may represent substantial differences in skill requirements. These differences may be reinforced by educational or licensing requirements. But despite the efforts that have gone into the construction of job ladders, there is little evidence that personnel management has succeeded in creating a system where skill distances between adjacent job ladders are equidistant. Sometimes these job ladders correspond more to conventional distinctions between jobs than to major skill differences. While relatively little is known about the skill differences among job ladders, some reasonable guesses can be made. Job-ladder shifts are probably more common for clerical jobs than for administrative or technical ones, and more common for administrative or technical jobs than for professional ones.

The most important type of job-ladder shift involves changing from a job ladder with a low ceiling to one with a high ceiling. So far as we know (Piore 1976; Althauser and Kalleberg 1981), most bureaucratized organizations divide

---

[1] In the federal civil service, this is called "career-conditional" status, which normally converts to career status after three years. When the term "tenured" is used in this paper, it will include both career and career-conditional status.

job ladders into rough tiers, with clerical and subprofessional ladders in a lower tier, and administrative and professional ladders in an upper tier. Advancement for lower-tier workers ultimately depends on their ability to cross the tier boundary. Organizations probably differ in the extent to which tier boundaries are porous, though little is known aside from case studies. (e.g., Northrup and Larson 1979; DiPrete and Soule 1986). However, the permeability of tiers probably varies within most organizations as well (e.g., DiPrete 1987). It should be easier to move from clerical to managerial positions than to move from subprofessional to professional ones, because the credential barrier is lower in the former case. The extent to which employees are able to use or are interested in using irregular positions as entry routes into the organization may also vary by level, or type of job ladder.

Promotion rates depend heavily on an organization's shape and demographic composition. But an individual's chances for promotion also depend on the management practices that handicap employees in organizational-promotion "tournaments" (Rosenbaum 1979). As we noted earlier, organizations generally favor tenured or tenure-track employees for promotions. The effects of education are generally strongest on the process of initial assignment. Otherwise, education is generally a more important factor in job-ladder switches than in movement up the rungs of a single ladder. Past performance also shapes a supervisor's appraisal of promotion potential. The advantages that allowed employees to move rapidly to their present position will generally continue to benefit them in the future. In contrast, a current supervisor may interpret an employee's slow progress in the past as indicating limited future potential (On the subject of age-graded mobility, also see Spilerman 1977; Kaufman and Spilerman 1982). An employee's particular route to a present position may also be related to further advancement. Managers have sometimes viewed certain types of clerical experience as disqualifying an employee for further advancement. Today, such attitudes have the potential to harm women more than men (Kanter 1977).

THE LOCATION OF GENDER-SPECIFIC BOTTLENECKS

Gender differences in advancement rates need not be uniform throughout the organizational hierarchy. On the contrary, the segmented character of job-ladder systems suggests that such uniformity is unlikely. Stewman and Konda (1983) have shown how an organization's shape and the seniority distribution of its employees can create promotion bottlenecks in certain grades. It is possible that gender-specific

bottlenecks exist as well. The correlation between gender and lower-tier position suggests that the barrier between the upper and lower tier may be one such bottleneck. At this grade range, lower- and upper-tier job ladders overlap. But women are more likely to be found on lower-tier job ladders. Near the tier boundary, therefore, a randomly selected woman should have a lower promotion rate than a randomly selected male, because she is more likely to be working on a lower-tier job ladder. At the lowest grades, in contrast, all job ladders have relatively low ceilings. One would not expect job ladder placement to matter as much there. Similarly, one would expect job-ladder placement to matter less in the higher grades than in the middle ones.

Group differences net of the structural opportunity offered by the job ladder can exist because of the association between gender and other resources or liabilities for promotion. These factors can also vary by grade level. At high grades, group differences may indicate management resistance to promoting the relatively few women workers found there to still more responsible positions.[2] Some researchers have found that informal networks help with promotion by teaching an employee the right way to act, by arranging sponsorship for promotion, and by other forms of peer support. Since men and women often have their own networks, men may have an advantage at organizational levels where males dominate (Lorber 1979; Miller and Garrison 1982; Roos and Reskin 1984). At low grades, net gender differences may indicate management preference for the relatively few men who work in female-dominated job ladders. At the middle grades, mobility for employees located near the grade ceiling of lower-tier job ladders occurs primarily through their ability to secure access to upper-tier jobs. Further mobility for administrative or professional trainees shows their success in passing through apprenticeship to achieve administrative or professional journeyman status. The transition from the female-dominated job world to the male-dominated job world may be the area where managerial discrimination and other disadvantages faced by women are most evident.

But disadvantages such as these do not guarantee that women will do worse than white males at all levels of the hierarchy. Workers from disadvantaged groups who reach high organizational levels have demonstrated that they can overcome gender- or race-specific

[2] President's Commission on the Status of Women 1963; see also Macy 1971, p. 85. A similar finding has also been reported for the private sector (see Wallace 1976).

disadvantages. Furthermore, equal-employment opportunity programs can partially or fully offset any disadvantages a worker faces because of gender. The precise character of gender differences in promotion rates must be assessed through empirical analysis.

DATA AND ANALYTIC STRATEGY

Personnel data for a one percent sample of the careers of white-collar employees of the federal government (Taylor and Grandjean 1977) active between 1972 and 1977 were used to examine gender differences in promotion rates in a government where women were generally in inferior positions (Taylor 1979; Grandjean 1981). Three distinct grade-promotion analyses were conducted. They covered the following ranges: (1) grade promotions in the lower grades (GS1-GS4); jobs at this level are clerical or technical and most workers are female; (2) grade promotions in the middle grades (GS5-GS10); here, higher clerical and technical positions as well as professional and administrative trainee positions can be found; and (3) grade promotions in the upper grades (GS11 and higher); these jobs are mostly professional and administrative, and they lie above the trainee grades.

In each of these analyses, the dependent variable measured whether a promotion of one or more grades occurred within two years of each first entry into a grade, to be described as the origin grade or position. All first entries occurred during the period between January 1972 and June 1975. The study focused on general schedule white-collar employees, which includes most white-collar employees outside the postal service.[3] While two years is a short time compared with a whole career, it is sufficiently long for a respectable proportion of employees to attain grade promotions and is appropriate for an analysis of this type.[4]

Promotion probabilities were specified as a function of the structural opportunity inherent in an employee's position and of individual-level factors pertinent to the personnel policies and practices that structure promotion tournaments. The following variables were used:

Ascribed Characteristics

1. FEMALE took the value 1 if female, and 0 if male.
2. MINORITY took the value 1 if nonwhite, and 0 if white.

Education, Military Service, and Work-experience

3. EDUCATION was measured by years of schooling and by a dummy variable, B.A. OR MORE, which took the value 1 if an employee had at least a bachelor's degree, and 0 if not.
4. VETERAN took the value 1 if the employee was a veteran, and 0 if not. In addition, an interaction between veteran's status and employment in the Department of Defense was included.
5. Two measures of the rapidity of career progress prior to reaching current grade were used. Estimated PREGOVERNMENT WORK EXPERIENCE was defined as age minus length of government service until the time of reaching the origin grade minus years of education minus 5. YEARS OF GOVERNMENT SERVICE was defined as the number of years of service prior to attaining the origin grade. Thus, an individual's age at reaching the origin grade is equal to 5 + education + estimated pregovernment work experience + government service. The separate coefficients for education, pregovernment work experience, and government service allow us to distin-

---

[3] Note that this analytic strategy, which was used for the grade-promotion analyses reported here, dictated that the base sample be limited to employees who entered grades in the time period under analysis. This sample is not representative of the total population of workers in these grades because employees who rarely changed grades are underrepresented. Similarly, a sample of employed individuals who became unemployed in a particular range of time could be used to study the factors that affect the rate of re-employment. But such a sample would not represent a random sample of people who were employed during the period under study.

[4] For technical reasons, we used a 25-month risk period. The analyses presented in this chapter used as the sample all instances of first entry into a grade during the period described above. Analyses in which the sample was restricted to employees who did not leave the government during the risk period, or to an employee's first entry into an origin grade during these 3.5 years,

were also performed. They gave substantially similar results. Promotions were measured through transactions posted to the computerized Central Personnel Data File (CPDF), the source of the data analyzed here. A check of the physical personnel files for a sample of employees showed that 20 percent of all transactions specifically reporting a grade promotion were missing from the CPDF, apparently at random (Nenni 1984). However, virtually all employees received an administrative pay increase each year. The transaction record for this event reported the employee's current grade, as would any other transaction record (reclassification, transfer, separation, temporary leave, and so forth) posted to an individual's personnel file. Therefore, while it was not always possible to determine the precise month of a grade promotion, it was possible to determine with acceptable accuracy whether a grade promotion occurred at some time during a 25-month interval.

guish the effects of each period on an employee's promotion chances.

Early Government Experience

6. FIRST GOVERNMENT JOB IRREGULAR indicated whether the employee had begun the career in an "irregular" status, i.e., in a temporary job, term job or part-time job. FIRST GOVERNMENT LOWER-TIER indicated whether the employee had begun the career in a lower-tier job (to be defined below).[5]

Structural Opportunity Associated with an Employee's Source Position

7. In the middle 1970s, white-collar jobs in the federal government were divided into approximately 450 separate job ladders. These ladders, which were officially called "occupational series," define "steps in the most natural line of promotion" (U.S. OPM 1981). For each job ladder, the MULTIPLE GRADE RATIO for each grade was computed from data found in the 1976 edition of *Occupations of Federal White-Collar Workers* (U.S. CSC 1976). Stewman and Konda's (1983) formula was modified to be job-ladder specific:

$$mgr_{il} = \sum_{j=i+1}^{K} \frac{n_{jl}}{n_{il}}$$

where l is the job ladder, i is the source grade, K is the top grade in the system (GS18), and n_{ij} is the number of employees located in grade j in ladder l in 1976.[6] This ladder-specific multiple-grade ratio cannot perfectly predict the ratio of vacancies to competitors in any given promotion tournament. The ratio of vacancies to competitors is also a function of the turnover rate, growth, recruitment patterns, job-ladder shifts, and heterogeneity in the shape of job ladders by agency or region. Nonetheless, we expected the ladder-specific multiple grade ratio to have an effect on outcomes.

8. THE PROPORTION FEMALE IN THE

JOB LADDER in 1976 (U.S. Civil Service Commission 1976) was included as an indicator of the extent to which a job ladder was sex-segregated.

Opportunities for Promotion Linked to Tenure Status and Other Measures of an Employee's Position

9. The origin grade.

10. Region and agency of the origin grade. These variables, while important, are not of compelling theoretical interest. They are not reported in the tables, to conserve space.

11. The dummy variable IRREGULAR STATUS indicated whether the origin position was temporary, term, or part-time.

12. The PATCO designation of a job indicated whether it was in the lower or upper tier and the type of job ladder within each tier. While employees on TECHNICAL and CLERICAL job ladders typically advanced one grade at a time, PROFESSIONAL and ADMINISTRATIVE job ladders were described as "two-grade interval" ladders, because incumbents who entered the trainee grades of 5 (the lowest), 7, or 9 typically advanced two grades at a time until they reached grade 11, which was usually the lowest journeyman position. Average grade and salary were much higher in the professional and administrative than in the technical and clerical job ladders. In 1976, for example, the average salary for professional, administrative, technical, and clerical employees was $23,187, $21,240, $13,708, and $12,237 (or $9,956 if postal clerks were excluded), respectively (U.S. CSC 1976). "MIXED" job ladders spanned the grade hierarchy and contained clerical, technical, and administrative positions, though individual agencies might make finer distinctions for these positions.

13. The dummy variable TEMPORARY LEAVE indicated any recorded temporary leaves of absence from the government during the two-year risk period for promotion.

Table 1 presents means and standard deviations for selected human-capital and structural variables for men and women in each of the three grade ranges in the samples under analysis.[7] At all three ranges, women had less education than men, though the difference was greatest in the middle grades, where the upper

---

[5] Because these two variables could be measured only for the portion of an employee's government work history between 1962 and 1977, they undercount irregular and lower-tier origins for employees who entered before then.

[6] In certain cases, the ladder-specific multiple grade ratio computed to an unreasonably large value. Since we did not expect its effect on outcomes to be linear over such a range, the measure of MGR actually employed in the chapter was min(mgr_{il}, 20). The value 20 was chosen after we investigated the distribution of the scores and their effects on the probability of promotion.

---

[7] Note that these means and standard deviations are not for all employees in the grade ranges indicated, but only those first entering grades in the period under study.

Table 1. Means and Standard Deviations of Selected Human Capital and Structural Variables, by Sex and Origin Grade

Variable	Grades 1–4		Grades 5–10		Grades 11–16	
	Male	Female	Male	Female	Male	Female
Years of education	13.0	12.7***	14.4	13.4***	15.8	15.5*
	(1.8)	(1.4)	(2.2)	(1.7)	(2.4)	(2.5)
B.A. or more	0.097	0.052**	0.44	0.18**	0.66	0.60*
Pre-govt work experience	6.6	9.3***	6.9	9.6***	6.2	6.3
	(8.4)	(10.0)	(7.7)	(9.4)	(7.3)	(7.5)
Years of govt service	3.5	2.6***	7.4	8.6***	12.1	12.1
	(3.9)	(3.4)	(7.2)	(7.2)	(8.6)	(9.3)
Proportion female in job ladder	0.42	0.80***	0.25	0.67***	0.14	0.32***
Multiple grade ratio	7.2	4.9***	6.6	2.8***	1.7	1.8
	(7.0)	(6.2)	(6.6)	(4.0)	(2.0)	(1.7)
Irregular status	0.37	0.24***	0.10	0.050***	0.025	0.031
N	1024	3145	2529	2308	2170	286

Note: Standard deviations are in parentheses.
* Significant at the 0.05 level.
** Significant at the 0.01 level.
*** Significant at the 0.001 level.

tier and the lower tier overlap. At the lowest and middle grades, it is perhaps surprising to find that more women than men had full-tenure (i.e., career or career-conditional) status. In general, the percent of employees with irregular status is much lower toward the top of the grade hierarchy. At all three levels, women were more likely to work in female-dominated job ladders than were men. The average percent female of women's job ladders ranged from a high of 79 percent in the lowest grades to 31 percent in the highest ones. Men also were more likely to work in female-dominated job ladders in the lower grades than in the higher ones, but, at all grade levels, the average percent female for men was lower than for women.

Table 1 also shows that average multiple grade ratio was higher for women working in the lower grades than in the middle or higher grades. The average MGR for men, in contrast, was nearly as high in the middle grades as in the lowest grades. The job ladders of women offered lower opportunities for advancement on average than did the job ladders of men in the middle grades, where men were more likely to be on upper-tier ladders than were women. But even in the lowest grades of the federal government, the MGR for men was much higher than for women. This reflected their greater tendency to work on technical ladders, which had higher grade ceilings than the clerical ladders where women worked. In the higher grades, however, women had the same level of structural opportunity for promotion as men, even though their job ladders were more heavily female. The percent female of one's job ladder may be correlated with opportunity overall, but the correlation is hardly perfect.

THE STRUCTURE OF PROMOTION IN LOWER GRADES

As Table 2 shows, women's zero-order promotion chances were equal to men's in the lowest grades. However, when controls were entered, women were at a net disadvantage. Education had only a slight effect on advancement in the lowest grades. Length of service mattered more; while the average length of service in the lowest grades was relatively short, those who had been there longer had lower promotion rates. Irregular status also clearly hurt one's chances for advancement, as hypothesized.

The ladder-specific multiple grade ratio had a positive and significant effect on promotion probabilities. As noted earlier, however, the multiple grade ratio did not completely account for structural opportunity. It should also be noted that job ladders did not constrain mobility very much at the lowest levels of government. Data not presented here showed that 14 percent of promotions from technical jobs, and a majority of promotions from clerical jobs, took the employee out of the old job ladder into a new one. At levels where job ladder boundaries, particularly clerical boundaries, were not very meaningful, the multiple grade ratio could not be expected to have strong predictive value. For this reason, the coefficient for technical jobs, which generally had higher grade ceilings than clerical jobs, was strongly positive.

Contrary to the usual expectations, the proportion female in the job ladder actually had a strong positive effect on advancement in the lower grades. This finding suggests that some structural force was creating vacancies in female-dominated job ladders at a greater rate

Table 2. Logistic Regression Coefficients for Grade Promotion from Grades 1–4

Coefficient	1	2	3	MEAN/SD
Intercept	0.053	−0.38	−0.98	—
	(0.4)	(0.4)	(0.05)	
Female	0.059	−0.39	−0.40	0.76
	(0.4)	(0.0005)	(0.0005)	
Minority	−0.054	−0.308	−0.16	0.28
	(0.4)	(0.6)	(0.05)	
Education (yrs)		0.039	0.062	12.8
		(0.2)	(0.04)	(1.5)
B.A. or more		−0.019	−0.10	0.061
		(0.9)	(0.6)	
Veteran		−0.29	−0.27	0.19
		(0.01)	(0.03)	
Veteran*DOD		0.25	0.36	0.052
		(0.1)	(0.06)	
Pre-govt work experience		−0.0030	−0.0016	8.6
		(0.4)	(0.7)	(9.7)
First govt job irregular		−0.40	−0.039	0.39
		(0.0000)	(0.7)	
Years of govt service		−0.043	−0.038	2.87
		(0.0001)	(0.0007)	(3.6)
Temporary leave		−0.034	−0.013	0.15
		(0.7)	(0.9)	
Grade 2		0.73	0.76	0.20
		(0.0002)	(0.0004)	
Grade 3		0.17	0.19	0.36
		(0.4)	(0.4)	
Grade 4		−0.27	−0.36	0.41
		(0.2)	(0.1)	
Proportion female in job ladder		0.75	1.24	0.71
		(0.0000)	(0.0000)	(0.28)
Multiple grade ratio			0.019	5.5
			(0.02)	(6.5)
Irregular status			−0.69	0.26
			(0.0000)	
Technical			0.76	0.11
			(0.0000)	
Other			0.25	0.15
			(0.05)	

Notes: 1. $N = 3950$, proportion promoted $(P) = 0.52$.
2. N, Means, and Standard Deviations apply to third column sample. Listwise deletion of missing values was used.
3. P-values and (in the last column) standard deviations are in parentheses.
4. The following coefficients were included in column 3 models in all tables, but are not reported, because of space considerations: Department of Defense, Agriculture, Commerce, General Services Administration, Health, Education and Welfare, Interior, Justice, Transportation, Treasury, Veterans Administration, and location of job (District of Columbia, eastern region, southern region, or western region).

than in male-dominated job ladders, net of what would be expected from the shape of these ladders (as measured by the ladder-specific multiple grade ratio). If turnover rates for women were greater than for men, turnover could provide such a force. But gender differences in turnover rates were not large enough to account for the finding. We believe that the proportion female was associated with higher rates of upward mobility at the very bottom grades because of a positive association

between the proportion female of a job ladder and the size of the net outflow from this ladder to other ladders. Many of the ladders with a high proportion of female workers had very low ceilings even for clerical job ladders. High outflow rates allowed workers on these ladders to continue advancing even after they reached the top of the ladder. In short, the ladder-specific MGR understated the promotional opportunities for such employees. But because the ceiling on their job ladders was so low, their

additional movement would still be in the clerical ranks. For example, further analysis showed that employees in such heavily female job ladders as clerk-typist or stenographer had significantly higher promotion rates in the grade 1–4 range when other factors (including, importantly, the MGR) were controlled. Summary data showed that 97 percent of all clerk-typists were in grades 1–4, and 98 percent of stenographers were in grades 1–5 (U.S. CSC 1976). Further advancement for such employees would require that they move to clerical ladders such as secretary, where 92 percent of the positions were in grades 5 and above. Net outflow from the top of ladders such as typist or stenographer would create room for advancement by workers in the bottom grades in these low-ceiling ladders. The net advantage for employees in female-dominated job ladders was not insignificant. Controlling for other factors, employees who would have a 0.4 probability of a grade promotion in two years would have a 0.49 probability if they were instead in a job ladder whose proportion female was 0.3 higher.

But the greater opportunity for advancement in female-dominated job ladders, controlling for other factors, should not obscure the net disadvantage of female employees compared with males. Under situations in which a man had a two-year promotion probability of 0.5, a comparable woman would have a predicted rate of only 0.4.[8]

THE STRUCTURE OF PROMOTION IN MIDDLE GRADES

In contrast to the zero-order equality in the promotion rates of women and men in the lowest grades, the overall promotion rate of women in the middle grades was much lower than that of men. While one-half of all males in grades 5–10 were able to obtain a grade promotion in two years, only about one-third of females were able to do so. However, the gender differences are largely explainable by the individual and organizational variables entered into the model in columns 2 and 3.

Women in the middle grades had less education than men (see Table 1). While education had little effect on mobility in the lowest grades, it had a strong positive association with promotion in the middle grades. This association remained even when organizational position was controlled. One reason that possession of a bachelor's degree continued to be significant in the presence of organizational

controls is that it eased the transition from the lower to the upper tier. We address this issue in more detail later.

Women's job ladders held lower prospects for advancement than men's at the middle levels, and this difference also explained part of the zero-order gender difference. The multiple grade ratio had a strong effect on job mobility at middle-grade levels, where a lower proportion of grade promotions involved job-ladder shifts (20 percent) than in the lowest grades (44 percent). The differences in opportunity offered by lower-tier and upper-tier jobs, or between female-dominated and male-dominated job ladders, was adequately captured by the MGR. Neither the ladder's PATCO status nor the proportion female in the job ladder was significant when this variable was included.

Table 1 showed that the MGR for women dropped substantially from the lower to the middle grades. The reason for this drop was the tendency for women to be on job ladders that had a ceiling in the middle grades. This slowdown can be demonstrated with the estimated coefficients. We computed predicted two-year promotion probabilities for a set of hypothetical employees. Each was a white female with 13 years of education, was not a veteran, had 8 years of pregovernment work experience, and worked on a job ladder that was 80 percent female.[9] They were located in grades GS-2 through GS-9, one to a grade. We assumed that the higher-graded employees had progressively more government seniority.[10] The predicted probability of promotion with two years of first entering a grade drops steadily from 0.73 to 0.59, 0.46, 0.35, 0.35, 0.36, 0.26 and 0.23, as she moved from grade GS-2 to grade GS-10. The predicted rate did not drop to zero near the job-ladder ceiling because, as we will see later, employees could move from their old ladder to a new one with a higher ceiling.

As in the lower grades, tenured employees at middle levels had higher promotion rates than irregular employees, though a lower proportion of workers at the middle levels were irregular. Employees who did not take a temporary leave during the risk period were also advantaged. Length of government service again had a strong negative effect on promotion rates, as did years of pregovernment work experience.

---

[8] The female coefficient was also significantly negative in analyses restricted to organizational survivors. This is no surprise, since their turnover rates were similar to those of males.

[9] Percent female rather than the MGR was used to simplify the computations.

[10] We assumed 1.5, 3, 4.5, 6. 8, 10, 12, and 14 years of seniority for grades 2, 3, 4, 5, 6, 7, 8, and 9, respectively. These assumptions are consistent with the mean values presented in Table 1. Column 3 coefficients from tables 2 and 3 were used.

Table 3. Logistic Regression Coefficients for Grade Promotion from Grades 5–10

Coefficient	1	2	3	Mean/SD
Intercept	0.053 (0.20)	0.73 (0.09)	0.28 (.06)	—
Female	−0.65 (0.0000)	0.0047 (0.96)	−0.20 (0.08)	0.51
Minority	−0.049 (0.5)	−0.025 (0.8)	−0.013 (0.9)	0.20
Education (yrs)		−0.0080 (0.8)	0.0054 (0.9)	14.0 (2.0)
B.A. or more		0.64 (0.0000)	0.39 (0.009)	0.33
Veteran		0.21 (0.04)	0.16 (0.2)	0.33
Veteran*DOD		−0.33 (0.009)	−0.0054 (0.97)	0.61
Pre-govt work experience		−0.031 (0.0000)	−0.025 (0.0000)	8.2 (8.8)
First govt job irregular		−0.25 (0.002)	−0.045 (0.6)	0.22
First govt job lower-tier		−0.032 (0.7)	0.057 (0.6)	0.54
Years of govt service		−0.056 (0.0000)	−0.044 (0.0000)	7.7 (7.1)
Temporary leave		−0.33 (0.003)	−0.25 (0.03)	0.11
Grade 6		0.14 (0.2)	0.059 (0.6)	0.13
Grade 7		0.29 (0.002)	0.16 (0.1)	0.25
Grade 8		−0.070 (0.7)	−0.26 (0.2)	0.043
Grade 9		−0.14 (0.2)	−0.25 (0.05)	0.21
Grade 10		−0.44 (0.09)	−0.85 (0.002)	0.022
Proportion female in job ladder		−0.81 (0.0000)	−0.27 (0.2)	0.45 (0.34)
Multiple grade ratio			0.074 (0.0000)	4.8 (6.0)
Irregular status			−0.43 (0.008)	0.072
Technical			0.11 (0.5)	0.24
Professional			0.10 (0.7)	0.13
Administrative			0.29 (0.1)	0.19
Mixed			0.084 (0.5)	0.18
Other			−0.57 (0.02)	0.08

$N = 4,114$, $P = 0.45$

Note: see notes at bottom of table 2.

THE STRUCTURE OF PROMOTION IN HIGHER GRADES

Compared with women in the lower or middle grades, the relatively few women who made it to the higher grades were similar to men, both in their attributes (see Table 1) and in their chances for promotion. Their chances were similar even though they worked on job ladders with significantly higher proportions female than did male workers. As we noted earlier, this fact was not associated with lowered structural opportunity; and so women were not disadvantaged in promotion tournaments at these levels of the civil service. But to repeat, relatively few women had reached these levels in the middle 1970s.

In summary then, the grade-promotion analyses reveal that women had difficulty advancing in the lower and middle grades. In the lowest grades, no zero-order differences existed between the promotion rates of women and men.

Table 4. Logistic Regression Coefficients for Grade Promotion from Grades 11–17

Coefficient	1	2	3	Mean/SD
Intercept	−1.13 (0.0000)	−0.25 (0.6)	−0.32 (0.6)	—
Female	0.18 (0.2)	0.12 (0.5)	−0.0021 (0.99)	0.12
Minority	0.12 (0.4)	0.17 (0.3)	0.11 (0.6)	0.097
Education (yrs)		0.057 (0.2)	0.046 (0.3)	15.8 (2.3)
B.A. or more		−0.22 (0.3)	0.018 (0.9)	0.66
Veteran		0.53 (0.0001)	0.27 (0.08)	0.55
Veteran*DOD		−0.68 (0.00001)	−0.015 (.96)	0.17
Pre-govt work experience		−0.047 (0.0000)	−0.046 (0.0000)	6.2 (7.4)
First govt job irregular		0.19 (0.2)	0.10 (0.5)	0.16
First govt job lower-tier		−0.37 (0.04)	−0.28 (0.2)	0.13
Years of govt service		−0.095 (0.0000)	−0.079 (0.0000)	11.9 (8.6)
Temporary leave		−0.44 (0.05)	−0.26 (0.3)	0.068
Grade 12		−0.53 0.32 (0.0000)	−0.49 (0.0006)	0.32
Grade 13		−1.08 (0.0000)	−1.06 (0.0000)	0.20
Grade 14		−1.74 (0.0000)	−1.91 (0.0000)	0.10
Grade 15		−2.52 (0.0000)	−3.29 (0.0000)	0.035
Grade 16		−0.64 (.4)	−1.05 (0.2)	0.0057
Proportion female in job ladder		0.15 (.7)	0.25 (0.5)	0.16 (0.18)
Multiple grade ratio			0.089 (0.003)	1.7 (2.0)
Professional			−0.18 (0.2)	0.43

$N = 2,112$, $P = 0.25$

Notes: No promotions in the sample occurred for GS-17 employees or irregular employees, who were omitted from the analysis. See notes at bottom of table 2.

But detailed controls for education, organiza-tional tenure status, and structural position revealed that women did somewhat worse in promotion tournaments than comparable men. In the middle grades, a strong zero-order gender difference existed, a difference largely due to gender differences in job ladder and education. Women and men had similar promotion rates in the higher grades, though women were under-represented at these levels. Clearly, organiza-tional location as well as personal characteristics contributed to gender differences in advance-ment rates. The structure of job ladders in the federal service concentrated these differences in a relatively narrow band of grades. This band spanned the boundary between the lower and upper tier of white-collar jobs.

CHANGE OF TIER, CHANGE OF TENURE

Tables 2, 3, and 4 make clear that opportunity for promotion depends on both job ladder and tenure status in the organization. Employees in the trainee grades of upper-tier job ladders had higher rates of promotion than comparably graded lower-tier workers. Employees with career or career-conditional status were pro-moted more rapidly than employees without such status. But neither job ladder nor tenure status is fixed. Both can change over time. Since such changes are an important mechanism behind upward mobility, it is important to look for gender differences in the rates at which workers changed these statuses. We performed two additional analyses. The first determined the probability of short-term promotion from lower-tier to upper-tier job ladders. The second determined the probability of promotion from temporary, term, or part-time to career or career-conditional tenure, which might be called a tenure promotion. Both analyses used the same set of variables as the grade-promotion analysis.

The base population for the tenure-promotion analysis was all employees who had irregular status in the period January 1972 through June 1975. The base population for the analysis of promotions from the lower-tier to the upper-tier population was all clerical and technical employ-ees. The Office of Personnel Management defines employees in mixed occupational series in grades 1–6 as clerical and in grades 7–10 as technical (U.S. OPM 1983), but some of these may actually have been moved through the ranks at the same rate as administrative trainees by their agencies. We limited the definition of lower-tier workers in mixed job ladders to employees in grades 1–4, plus those in grades 5 or 6 who already had served more than three years in the federal civil service, and thus could not have been newly hired administrative

trainees. The upper-tier destination positions were defined to include all exclusively adminis-trative and professional job ladders. Due to the ambiguous status of positions on mixed job ladders in the middle grades (which by official convention would be classified as technical, not administrative, positions in any case), they were excluded from the definition of the upper tier.[11]

Table 5 shows that approximately eight percent of the lower-tier workers were able to obtain promotions to professional or exclusively administrative job ladders during a two-year period. While eight percent is not a large figure, two years is not a long time, either. Further-more, the rate of mobility to the upper tier was higher for lower-tier employees located in higher grades, as can be seen by the size of the grade coefficient in Table 5. Earlier in this paper, we described the estimated rate of movement for a set of hypothetical lower-level female employees. If we use the coefficients from Table 5 to compute the predicted probabil-ity that our hypothetical employee would shift from the lower tier to a professional or exclusively administrative job ladder in a two-year period as a function of her grade, the results for grades 5 through 9 would be 0.08, 0.09, 0.10, 0.11, and 0.13, respectively. In addition, as we noted above, employees could shift to mixed job ladders, which also would offer prospects for further advancement, if not necessarily at the same rate as positions on the two-grade interval administrative or professional job ladders. In short, the boundary between the lower and the upper tier was partially permeable during these years in the federal civil service.

However, the boundary was not so permeable as to make job ladders meaningless. As we have already seen, the rate of grade promotion at middle levels clearly depended on the character of the job ladder (see Table 3), and slowed down near job-ladder ceilings. Thus, the observed rate of promotion from the lower to the upper tier did not eliminate the disadvantage of low job ceilings. Furthermore, it did not

---

[11] The ambiguous character of mixed job ladders is demonstrated by the 1979 split of the mixed job ladder 301, "general clerical and administrative," into two job ladders: 301, "miscellaneous administrative and pro-grams," which was defined as an administrative job ladder; and 303, "miscellaneous clerk and assistant," which was defined as a mixed clerical-technical job ladder. In 1976, 28,000 employees on job ladder 301 were located in grade 5 and 17,000 in grade 7. In 1983, 1,926 employees were located in GS-5 on job ladder 301, now an exclusively administrative job ladder; 14,474 were located in GS-5 on 303; 2,827 were located in GS-7 on 301; and 8,765 in GS-7 on 303. It appears that most of the positions in grades 5–8 of 301 in the years preceding its division were in fact considered to be clerical-technical jobs.

Table 5. Logistic Regression Coefficients for Promotions from Lower to Upper Tier and from Noncitizenship to Citizenship Status[†]

Coefficient	Change of Tier		Change of Tenure	
	1	3	1	3
Intercept	−2.08 (0.0000)	−3.94 (0.0000)	−0.21 (0.009)	1.03 (0.2)
Female	−0.53 (0.0000)	−0.66 (0.0000)	0.40 (0.0001)	−0.21 (0.2)
Minority	0.051 (0.6)	−0.025 (0.8)	0.23 (0.07)	0.30 (0.05)
Education (yrs)		0.043 (0.3)		−0.042 (0.4)
B.A. or more		0.97 (0.0000)		0.0035 (0.99)
Veteran		0.075 (0.7)		0.21 (0.3)
Veteran*DOD		0.063 (0.8)		−0.19 (0.6)
Pre-govt work experience		−0.019 (0.001)		−0.0040 (0.6)
Years of govt service		−0.043 (0.0000)		−0.017 (0.2)
Temporary leave		−0.85 (0.0001)		−0.66 (0.0002)
Grade		0.21 (0.0000)		0.052 (0.2)
Proportion female in job ladder		1.24 (0.0001)		1.10 (0.002)
Multiple grade ratio		−0.042 (0.002)		−0.025 (0.02)
Irregular status		−0.065 (0.7)		
Technical		−0.11 (0.6)		−0.23 (0.4)
Mixed		1.1 (0.0000)		−0.40 (0.03)

Notes: 1. For promotion from lower to upper tier, $N = 6,457$, $P = 0.081$.   [†] = See notes at bottom of table 2.
2. For change of tenure status analysis, $N = 1,398$, $P = 0.52$.

eliminate the group disadvantage women experienced because they were concentrated in lower-tier jobs (see the first column of Table 3). Finally, Table 5 shows that women were at a significant zero-order disadvantage in obtaining status promotions from the lower to the upper tier. This advantage could not be explained by the introduction of control variables. Even when education, grade, type of job ladder, the proportion female in the job ladder, and the multiple grade ratio were controlled, women had significantly lower probabilities of shifting from lower-tier to upper-tier job ladders than did men.

But while women were at a significant disadvantage in gaining status promotions, employees who worked on job ladders with a high proportion female were actually at an advantage, when other factors were controlled. The advantage was not huge (a 0.3 increase in proportion female lifted the predicted probabil-ity of promotion in a two-year period from 0.10 to 0.14), but it was highly significant. This association occurred because shifts from lower-tier to upper-tier positions were rare from technical jobs in engineering and the physical sciences, where few women worked. They were comparatively frequent in technical jobs in law (claims examiners of various sorts), where women were found in much greater numbers. When controls were entered for a select set of occupational groups, the coefficient for proportion female became small and statistically insignificant.[12]

It should be pointed out, though, that employees in the male-dominated technical positions in engineering and the physical sciences had relatively high grade ceilings

___

[12] The addition of these occupational controls did not materially affect the gender coefficient ($b = -0.60$, $p = 0.0001$).

compared with most lower-tier jobs. The average salary of engineer technicians was $17,402 in 1976 (U.S. CSC 1976), while physical science technicians made an average of $14,766, compared with the $13,703 average for technicians generally, and the $9,976 average for clerical workers outside the postal system. The greater opportunity for grade advancement in these jobs than in most lower-tier positions offset the disadvantage of a lower net rate of promotion to upper-tier job ladders.

In sum, the employees who found it easiest to shift from the lower to the upper tier were male and worked in the higher grades of the lower tier.[13] They had college degrees[14] and relatively short government careers to date. They did office work as opposed to scientific work and (judging from the sign of the MGR coefficient) had nearly reached the limit for advancement offered by their current job ladder. To put it another way, the lower-tier employees who found it easiest to make the switch were those who were most similar (in age, education, and gender) to those newly hired into upper-tier job ladders. Since men were at a significant advantage in gaining promotions from the lower to the upper tier, their share of status promotions was disproportionate to their representation in the lower tier population. While 68 percent of the base population was female, women obtained only about 57 percent of the promotions.

Rates of tenure promotion were much higher than rates of promotion from the lower to the upper tier. Table 5 shows that over 50 percent of the irregular employees were able to change their tenure status during a two-year period of time. Women had a substantial zero-order advantage in making this switch, but this difference disappeared in the presence of controls. As hypothesized, irregular status was an important entry route into clerical, female-dominated job ladders. While irregular employees had difficulty achieving grade promotions compared to career employees, they found it comparatively easy to make the transition to tenured status. Moreover, as Tables 2 and 3 show, they suffered no significant handicap once this transition had been made, if compared

---

[13] Since promotions from the lower tier to the upper tier were comparatively rare (though not impossible) in the lowest grades of the lower tier, we performed another analysis that restricted the sample to lower-tier employees in grades 4 and higher. The results still showed women to have a highly significant ($b=0.56$, $p=0.002$) disadvantage in obtaining promotions to the upper tier.

[14] Ten percent of lower-tier employees in grades 4 and higher had college degrees.

with similarly situated workers who entered the government as full citizens.

DISCUSSION

Although many studies have demonstrated that women do not achieve the same rate of career growth, either in status or income as do men, studies based on national samples cannot easily pinpoint the reasons for these differences. We have argued that gender differences in career advancement can be due to sex segregation, to tenure status, to unequal opportunity for changing job ladder, or to management practices that disadvantage women for other reasons, such as discrimination. This study focused on promotion within a highly bureaucratized labor market. It showed that the use of organizational data can disentangle the processes that potentially inhibit the advancement of women.

Clearly, all women in the federal government were not in the same boat in the middle 1970s. Those who had reached administrative or professional journeyman levels (five percent of women in the current analysis) enjoyed the same rate of subsequent advancement as did men. But women experienced three difficulties reaching journeyman status: The first involved recruitment. Although not the subject of this paper, women generally entered the federal hierarchy at lower grade levels than men.[15] In a sex-neutral promotion system, such differentials would tend to persist over time. Second, women tended to advance in the crucial middle grades at lower rates than men. Most of this difference was due to the concentration of women in female-dominated lower-tier job ladders, which offered relatively little opportunity for advancement. Third, while it was not impossible to obtain promotions from lower-tier to upper-tier job ladders, it was harder for women to do so than for men, even when other factors were controlled.

Our analysis showed that there is no necessary relationship between sex segregation and career advancement. What matters is whether women's jobs offer advancement prospects comparable to those of men. The relatively few women who had achieved journeymen status in the government worked on job ladders that had a relatively (for these levels) high female composition. But nonetheless, these ladders were structured to offer the same opportunity for future advancement that men enjoyed. The result was gender equality in promotion rates. As a practical matter, however, good jobs that are not male-dominated are still hard to find in the

---

[15] Supporting data are available on request from the authors.

nation's labor market, though women's representation in certain managerial and professional occupations has increased in recent years (Rytina and Bianchi 1984; England and Farkas 1986).

Our results also suggest that the disadvantage of working on a low-ceiling job ladder go beyond the problem of gaining a promotion in the short-term. At any given grade, older employees were at a disadvantage in competing for promotions. Furthermore, this disadvantage tended to increase with organizational level. In other words, factors that slow an employee's present advancement may contribute to a developing handicap that lowers an employee's ultimate career attainment. Since sex-segregation* is typically associated with unequal opportunity for promotion within the job ladder, it presents a long-term as well as a short-term problem for women interested in successful careers.

Finally, our analysis showed that the form of the employment relationship can affect promotion rates. Employees with irregular tenure status advanced more slowly than career employees. In the federal workforce of the 1970s, this fact did not contribute to overall gender differences in promotion rates. But in the labor market as a whole female workers more often work part-time, and have lower advancement rates as a consequence (Corcoran, Duncan, and Ponza 1984). Thus, both differences in employment status and differences in job placement can work to the disadvantage of women.

The implications of our results for policy are clear. Sex-neutral promotion policies will obviously not correct gender inequality as long as women and men are hired into different levels of an organization. If women and men are hired into organizations at similar grades, gender-differences in advancement will still exist as long as the opportunities for advancement up the job ladder are lower in women's job ladders than in men's. Gender differences that persist even in the presence of controls for job ladder are most prominent in the process of transition from lower- to upper-tier jobs. Women in upper-tier jobs may still find themselves disadvantaged relative to men in the competition for promotions, but their disadvantage is probably smaller than that of lower-ranking women.

Whether the results from this study generalize to other bureaucratized organizations cannot be definitively answered here. However, the federal personnel system is similar to that employed in large bureaucratized firms. The use of job ladders and formal personnel procedures for promotion decisions is common to government and business bureaucracies (Pfeffer and Cohen 1984, Baron et al. 1986). Therefore, many of our results are probably applicable to the private sector as well as to other civil service systems. Further study is needed to assess the generality of these results and to understand why the transition from lower- to upper-tier jobs is so difficult for women even in a bureaucratic labor market.

REFERENCES

Althauser, Robert P. 1987. "Internal Labor Markets U.S.A.: A Thematic Review." Paper presented at the 1987 Annual Meetings of the American Sociological Association.

Althauser, Robert P. and Arne L. Kalleberg, 1981. "Firms, Occupations and the Structure of Labor Markets: A Conceptual Analysis." Pp. 119–49 in *Sociological Perspectives on Labor Markets*, edited by Ivar Berg. New York: Academic.

Baron, James N., Alison Davis-Black, and William T. Bielby. 1986. "The Structure of Opportunity: How Promotion Ladders Vary within and among Organizations." *Administrative Science Quarterly* 31:248–73.

Beller, Andrea H. 1982. "Occupational Segregation by Sex: Determinants and Changes." *Journal of Human Resources* 17:371–92.

Bielby, William T. and James N. Baron. 1984. "A Woman's Place is with other Women: Sex Segregation within Organizations." Pp. 27–55 in *Sex Segregation in the Workplace: Trends, Explanations, Remedies*, edited by Barbara Reskin. Washington, DC: National Academy Press.

Burstein, Paul, 1985. *Discrimination, Jobs and Politics: The Struggle for Equal Employment Opportunity in the United States since the New Deal*. Chicago: University of Chicago Press.

Corcoran, Mary, Greg J. Duncan, and Michael Ponza. 1984. "Work Experience, Job Segregation, and Wages." In *Sex Segregation in the Workplace: Trends, Explanations, Remedies*, edited by Barbara Reskin. Washington, DC: National Academy Press.

DiPrete, Thomas A. 1987. "The Professionalization of Administration and Equal Employment Opportunity in the U.S. Federal Government." *American Journal of Sociology* 93:119–40.

DiPrete, Thomas A. and Whitman T. Soule. 1986. "The Organization of Career Lines: Equal Employment Opportunity and Status Advancement in a Federal Bureaucracy." *American Sociological Review* 51:295–309.

Doeringer, Peter B. and Michael J. Piore. 1971. *Internal Labor Markets and Manpower Analysis*. Lexington, MA: Heath.

England, Paula and George Farkas. 1986. *Households, Employment, and Gender: A Social, Economic, and Demographic View*. NY: Aldine.

Grandjean, Burke D. 1981. "History and Career in a Bureaucratic Labor Market" *American Journal of Sociology* 86:1057–92.

Kanter, Rosabeth M. 1977. *Men and Women of the Corporation*. New York: Basic Books.

Kaufman, Robert L. and Seymour Spilerman, 1982. "The Age Structure of Occupations and Jobs." *American Journal of Sociology* 87:827–51.

Lorber, J. 1979. "Trust, Loyalty, and the Place of Women in the Informal Organization of Work." Pp.

371–81 in *Women: A Feminist Perspective*, edited by
J. Freeman, Palo Alto, CA: Mayfield.

Macy, John, 1971. *Public Service: The Human Side of
Government*. New York: Harper and Row.

Miller, Joanne and Howard H. Garrison. 1982. "Sex
Roles: The Division of Labor at Home and in the
Workplace." *Annual Review of Sociology* 8:237–62.

Nenni, Ralph. 1984. "Initial Quality Assessment of
Federal Personnel Statistics Program Ten Percent
Sample (1962–72) and Central Personnel Data File
(CPDF) Transactions History File (1973–1980)." U.S.
Office of Personnel Management Compliance and
Investigations Group. Unpublished paper.

Northrup, Herbert R. and John A. Larsen. 1979. *The
Impact of the AT&T-EEO Consent Decree*. Philadel-
phia: University of Pennsylvania Press.

O'Farrell, Brigid and Sharon Harlan. 1984. "Job
Integration Strategies: Today's Programs and Tomor-
row's Needs." In *Sex Segregation in the Workplace,
Trends, Explanations, Remedies*, edited by Barbara
Reskin. Washington, DC: National Academy Press.

Pfeffer, Jeffrey and Yinon Cohen, 1984. "Determinants
of Internal Labor Market Arrangements in Organiza-
tions." *American Sociological Review* 29:550–72.

Piore, Michael J. 1976. "Notes for a Theory of Labor
Market Stratification." Pp. 125–50 in *Labor Market
Segmentation*, edited by Richard C. Edwards, Michael
Reich, and David M. Gordon. Lexington, MA:
Lexington Books.

President's Commission on the status of Women. 1963.
*Report of the President's Commission of the Status of
Women*. Washington, DC: U.S. Government Printing
Office.

Roos, Patricia A. 1985. *Gender and Work: A Compara-
tive Analysis of Industrial Societies*. New York: State
University of New York Press.

Roos, Patricia A. and Barbara F. Reskin. 1984.
"Institutional Factors Contributing to Sex Segregation
in the Workplace." In *Sex Segregation in the
Workplace: Trends, Explanations, Remedies*, edited by
Barbara Reskin. Washington, DC: National Academy
Press.

Rosenbaum, James E. 1979. "Tournament Mobility:
Career Patterns in a Corporation." *Administrative
Science Quarterly* 24:220–41.

Rosenfeld, Rachel. 1980. "Race and Sex in Career
Dynamics." *American Sociological Review* 45:583–609.

———. 1983. "Sex Segregation and Sectors: An
Analysis of Gender Differences in Returns from
Employer Changes." *American Sociological Review*
48:637–65.

Rytina, Nancy F. and Susan Bianchi. 1984. "Occupa-
tional Reclassification and Distribution by Gender."
Monthly Labor Review 107(3):11–17.

Spilerman, Seymour. 1977. "Careers, Labor Market
Structure, and Socioeconomic Achievement." *Ameri-
can Journal of Sociology* 83:551–93.

Stewman, Shelby and Suresh L. Konda. 1983. "Careers
and Organizational Labor Markets: Demographic
Models of Organizational Behavior." *American Jour-
nal of Sociology* 88:637–85.

Taylor, Patricia A. 1979. "Income Inequality in the
Federal Civilian Government." *American Sociological
Review* 44:468–79.

Taylor, Patricia A. and Burke D. Grandjean, 1977. One
Percent Extract from the Central Personnel Data File
and 10% Extract from the Federal Personnel Statistical
Program, obtained through funding by the National
Institute of Education.

U.S. Civil Service Commission. 1976. *Occupations of
Federal White Collar Workers*. Washington, DC: U.S.
Government Printing Office.

U.S. Office of Personnel Management. 1981. *Handbook
of Occupational Groups and Series of Classes*.
Washington, DC: U.S. Government Printing Office.

———. 1983. *Occupations of White-Collar and Blue-Col-
lar Workers*. Washington, DC: U.S. Government
Printing Office.

Wallace, Phyllis A. 1976. *Equal Employment Opportu-
nity and the AT&T Case*. Cambridge, MA: MIT Press.

Wolf, Wendy and Neil D. Fligstein, 1979. "Sex and
Authority in the Workplace." *American Sociological
Review* 44:235–52.

RELIGIOUS ECONOMIES AND SACRED CANOPIES: RELIGIOUS MOBILIZATION IN AMERICAN CITIES, 1906*

ROGER FINKE
Loyola University of Chicago

RODNEY STARK
University of Washington

For generations, sociologists have believed that cities are less hospitable to religion than are rural areas and that where many faiths compete for followers, the credibility of each is reduced. In this essay we attempt to explain why these received truths are, in fact, nostalgic myths. We try to demonstrate that religious participation is and ought to be higher in cities and that competition among religious bodies increases levels of religious mobilization. Our analysis is based on the 1906 U.S. Census of Religious Bodies, and the units of analysis are the 150 largest cities—all of those having an estimated population of 25,000 or more.

INTRODUCTION

Since the dawn of recorded history, the city has been depicted as a precinct of sin and impiety. Whether the cities were Sodom and Gomorrah, Corinth and Rome, or New York and Chicago, observers took it for granted that city life not only encourages vice, but also fosters skepticism and doubt—the pious life is easier in the hinterlands. John Lancaster Spalding wrote in 1880: "In the city neither the rich nor the poor can realize the infinite charm of the Christian ideal. The heart is troubled there, and God is not in the whirlwind of human passion" (Spalding [1880] 1967, p. 8).

Not only prophets and preachers regard the city as wicked and secular. Social scientists also assert that city life is corrosive of the moral order and that urban life is inevitably more permissive than country life. As Gideon Sjoberg (1960, p. 340) explained, urban religion sustains norms that "are generally permissive" because the "divergent, and often contradictory, roles, and the new technology ensures a continuous cycle of change, all of which requires flexibility of norms." Moreover, from Durkheim on, sociologists have been certain that the unavoidable tendency towards religious pluralism in cities weakens faith. Indeed, Harvey Cox (1965, p. 1) flatly asserted that secularization occurred because the "cosmopolitan confrontations of city living exposed the relativity of the myths and traditions men once thought were unquestionable." Similar views inspired Peter Berger's book, *The Sacred Canopy* (1967), one of the more influential recent works in the sociology of religion.

In this essay, we suggest that the received wisdom about the relationship between cities and religion is a nostalgic myth. We show that urbanites are far more likely than rurals to *actively participate* in religion and that pluralism causes levels of activity and participation *to increase*.

To this end, we examine the religious impact of urbanization in the United States. We also examine the link between the religious pluralism of communities and their degree of religious mobilization. In pursuit of these matters, we journey back to the turn of the century, to a time when urbanization was rapid and the religious diversity of American cities was expanding. But, first let's clarify the central issues in dispute.

RELIGIOUS PLURALISM

It is not surprising that modern sociologists expect the religious pluralism of the city to lead to the decline of religion. In his famous treatise on suicide, Durkheim ([1897] 1951) condemned urban pluralism as both cause and consequence of the breakdown of moral integration. Where multiple religious groups compete, each discredits the other and encourages the view that religion per se is open to question, dispute, and doubt. Durkheim believed that, in pluralistic societies where there are multiple religious options, people are cast adrift in a sea of moral uncertainty, which, in turn, produces all manner of social pathologies.

In similar fashion, Berger (1967) argued that pluralism fractures the "sacred canopy" of a society—a canopy that exists only when all (or nearly all) members of a society assent to "One True Faith." Although Berger is speaking of religious plausibility structures and not religious institutions, he does not exclude the institutional church from this religious crisis. Indeed, in his more recent book, *The Heretical Imperative*

(1979), Berger clearly states his position: "Modernity has plunged religion into a very specific crisis, characterized by secularity, to be sure, but characterized more importantly by pluralism. In the pluralistic situation, . . . the authority of all religious traditions tends to be undermined (1979, p. xi)." Pluralism forces religious plausibility structures to compete and deprives all religions "of their status as taken-for-granted" (1967, p. 151). For Berger, pluralism forces religions to compete, and this competition forces all religions into a market situation, a situation he views as threatening to the future of religion.

We agree with Berger that pluralism forces religions to compete for adherents. Unlike Berger, however, we view competition as a stimulus for religious growth and not an avenue for its demise.

RELIGIOUS ECONOMIES

Berger is one of very few social scientists to apply an explicit market model to religious organizations. The immense advantage of this approach is that it permits examination of the extent to which particular religious organizations are shaped by external imperatives.[1] In pursuit of powerful theoretical propositions about religious movements, we have developed the notion of the *religious economy* (Stark 1985; Stark and Bainbridge 1985, 1987; Finke and Stark 1986; Stark and Finke forthcoming).

Religious economies are like commercial economies. They consist of a market and a set of firms seeking to serve that market. Like all market economies, a major consideration is their degree of regulation. Some religious economies are virtually unregulated, while others are restricted to state-imposed monopolies. Deductions from a general theory of religion (Stark and Bainbridge 1987) suggest that, to the degree a religious market is unregulated, pluralism will thrive. That is, the "natural" state of religious economies is one in which a variety of religious groups successfully cater to the special interests of specific market segments. This arises because of the inherent inability of a single religious organization to be at once worldly and other-worldly, while the market will always contain distinct consumer segments seeking more and less worldly versions of faith. Indeed, because of this underlying differentiation of consumer preferences, religious economies never can be successfully monopolized, even when a reli-

gious organization is backed by the state. Even at the height of its temporal power, the medieval church was surrounded by heresy and dissent (Johnson 1976; Stark and Bainbridge 1985). Of course, when repression is great, religions competing with the state-sponsored monopoly will be forced to operate underground. But whenever and wherever repression falters, lush pluralism breaks through.

We part company with Berger on the consequences of outbreaks of pluralism. Along with most sociologists, he believes in the superior organizational power of monopoly faiths: by providing the people with a single plausibility structure, the monopoly religion can inspire the kind of deep faith we often associate with the medieval village. Here is precisely where we think the nostalgic errors enter. There is ample evidence that in societies with at least a putative monopoly faith, religious indifference is rife. Our contrary conceptions of religious commitment in earlier eras are simply wrong. In the Puritan Commonwealth of Massachusetts on the eve of the Revolution, the rate of religious adherence was about 16 percent, compared with better than 60 percent today (Stark and Finke forthcoming). Historians of the medieval church now recognize that huge areas often were nearly untouched by church influence (Johnson 1976). And today, close inspection of the religious situation in most nations where "everyone" is a Roman Catholic reveals levels of religious participation that are astoundingly low compared with the United States and Canada. Ireland, Poland, and Quebec have long been exceptions to this pattern. However, in each instance the church has served as the primary institution of political resistance to external domination. In effect, religious participation in these societies proves commitment to nationalism.

The inability of the monopoly church to mobilize massive commitment is inherent in the segmentation of any religious market. The fact is that a single faith cannot shape its appeal to precisely suit the needs of one market segment, without sacrificing its appeal to another. In contrast, where many faiths function within a religious economy, a high degree of specialization occurs. It follows that many religious bodies will, together, be able to meet the demands of a much larger proportion of a population than can be the case where only one or very few faiths have free access. Moreover, in faith as in finance, monopoly firms will be lazy. Since religious monopolies can only exist as creatures of state power, monopoly faiths are not exposed to market forces. In contrast, in an unregulated religious economy, faiths seek to maximize their efforts to attract and to hold members—those that can't compete will disap-

---

[1] The literature on religious movements is dominated by case studies that stress the importance of *internal* factors in shaping events. Viewed in context, however, it often is clear that the fate of many groups is determined almost entirely by external factors beyond their control.

pear. *The more pluralism, the greater the religious mobilization of the population—the more people there are who will be committed to a faith.* Put another way, the more highly specialized and aggressive the churches are, the greater the odds that any given individual will be activated.

Some will object that this is participation, not faith, and that full and active churches are not the same thing as a society in which all partake of a unitary sacred canopy. But that line of thought implies that religion is stronger where it is neglected than where members are willing to commit much time, money, and attention.

Our disagreement with the traditional sociological position on the impact of urbanism and pluralism on religion is summed up by the following propositions. Other things being equal, (1) Urbanization increases levels of religious mobilization; and (2) Pluralism increases levels of religious mobilization.

THE DATA

Unlike today, religious statistics for American cities were plentiful at the turn of the century. The U.S. Census Office (1910) compiled a special report entitled *Religious Bodies: 1906* that reported on 14 items of information about religious bodies and their individual churches.[2] These items included such gems of information as the number of members by sex, the year the church was established, the number of Sunday schools and Sunday school teachers, the language in which services were conducted, as well as the number of and salary of ministers. Many of these items were summarized by denomination and broken down by county, state, and the major cities of the United States.

Although it has seldom been used for research, the 1906 special report is one of the most complete censuses ever conducted on the religious bodies of America. It provides religious statistics on 186 denominations, as well as numerous independent churches throughout the United States—an increase of 41 denominations over the 1890 report. Of these 186 denominations, the census reports that "three bodies made

a full report, while the majority made a report for from 99.2 to 99.9 per cent of the total number of organizations (U.S. Census 1910, p. 24)." Only 1 of the 186 religious bodies reported on less than 90 percent of their organizations; only 65.1 percent of the Jewish congregations gave a full report. In addition to its very thorough and effective methods for collecting the data, the Census Bureau reported making great efforts to verify the data. "The returns . . . were carefully compared with the yearbooks and minutes of the various bodies, and with all other available sources of information, and supplemental information was obtained by special correspondence, wherever this was found necessary (U.S. Census 1910, p. 11)." The Census Bureau concluded that its religion statistics were "very thorough and complete."

We make only one revision to the statistics the census provides: we *standardize* the definition of membership across denominations. Since some denominations count children as members (e.g., Catholics and Lutherans) while others do not (e.g., Baptists and Methodists), we base our statistics on the *total number of adherents* to a religious group—a measure that includes children as well as adults (Johnson, Picard, and Quinn 1974). For denominations that did not count children, we simply inflated their reported membership by factoring in the local ratio of children 13 and under to the total population.[3] The result is a standardized measure of church adherents, or an *adherence rate,* for all denominations. These rates, based on the excellent statistics of the *Religious Bodies: 1906,* give us an opportunity to look at the level of religious mobilization, as well as the extent of religious pluralism, of major American cities at the start of the 20th century.

URBANIZATION AND
RELIGIOUS MOBILIZATION

At the turn of the century, American cities had experienced several decades of explosive growth. Some of this growth was the result of immigration, but there also was a growing torrent of migrants from rural America. Did these migrants leave their church affiliation behind as they fled the quiet piety of rural life for the worldliness of the city?

Data in Table 1 reveal that religious participation was *higher* in the cities than in the surrounding small villages and hinterlands. For cities with a population of 25,000 or more, the adherence rate was six percentage points higher

---

[2] A complete listing of the information collected from each church is as follows: (1) Denomination; (2) Division (ecclesiastical); (3) Organization; (4) Location (city, town, or village; county; state); (5) Year in which established; (6) Number of church edifices; (7) Seating capacity; (8) Value of church property; (9) Amount of debt on church; (10) Value of parsonage, if any; (11) Language in which services are conducted; (12) Ministers (number of; salary); (13) Communicants or members (total number; males; females); (14) Sunday schools conducted by church organization (number of schools; number of officers and teachers; number of scholars).

[3] The exact formula used to convert adult membership into total adherents is: (Adult Members * (Total Population / (Total Population—Children 13 and under).

Table 1. Percent Church Adherents for Cities of over 25,000, for Rural Areas, and for the Nation

	% of Population Church Adherents
All cities 25,000 + (N = 150)	56
Cities 25,000–50,000 (N = 65)	60
Cities 50,000 + (N = 85)	55
Rural areas (nation minus cities)	50
National rate	51

than for the remainder of the nation. The rate for small cities (25,000 to 50,000) might be slightly inflated, since people from the surrounding areas might attend church in the city, but even for the large cities (50,000 +), the rate of adherents was higher than for the nation as a whole. Cities, in general, had a higher rate of religious involvement than the surrounding small towns and the hinterland.

This result may appear surprising, but it can readily be explained. From a practical standpoint, it was much easier to attend church in an urban area than in a rural area. All cities had churches, but the lower population density of rural areas often made it hard to sustain churches. Even when there were rural churches, people often had to travel relatively long distances to reach them. We also must keep in mind the degree to which rural residents formed small, relatively closed networks impervious to social pressures or outside attachments. These closed networks often sustained various forms of quiet deviance, lack of religion being among the least serious. An additional factor is, of course, pluralism. Americans in cities always had a much greater range of available choices, and urbanites have always been exposed to more intensive recruitment efforts.

A third aspect of the urban community that might have increased religious participation was Catholicism. At the turn of the century, the Catholic Church had become the largest single Christian body in the nation, was undergoing the most rapid growth (primarily through immigration), and was centered in the major urban areas of the nation. In this setting, the Catholic Church not only was discovering its ability to compete in a religious economy, it added a new competitive intensity to the religious market.

Table 2. Percent Roman Catholic for Cities of over 25,000, for Rural Areas, and for the Nation

	% of Population Catholic	% of Adherents Catholic
All cities 25,000 + (N = 150)	33	60
Rural areas (nation minus cities)	11	22
National rate	17	32

fear and prejudice. In the United States of the late 19th and early 20th century, Catholics weren't just another Christian firm active in the marketplace. Seen through many Protestant eyes, they were the dreaded monopolists, the bloody foes of the Reformation and a potential threat to a Protestant status quo. As Catholicism grew, Protestantism grew anxious. By 1906, the Catholic segment had become large throughout the nation, but it was most dominant in the cities (see Table 2). Approximately one third of all urban residents were Catholic. Catholicism had become a prominent force in the religious market of urban America.

THE EFFECTS OF PLURALISM

To measure the level of religious pluralism, we have used a religious diversity index that accounts for both the *number* of different denominations and *size* of each denomination in a given city. This index is based on a probability equation commonly used to measure linguistic diversity (Greenberg 1956; Lieberson 1964). The basic equation is:

$$1 - ((a/z)^2 + (b/z)^2 + (c/z)^2 + (d/z)^2 \ldots)$$

Where a, b, c, d, and so forth, each represents the number of adherents in a particular denomination and z represents the total number of adherents. Therefore, if all adherents of a given city were from one denomination the diversity score would be zero, but if they were spread evenly across the 98 major denominations used to compute the diversity index, the score would approach unity. As it turned out, the actual range of scores was between .27 and .90.[4] This index provides a summary measure of religious pluralism for each of the major cities in 1906.

---

[4] The number of denominational categories included in the equation was reduced from the original 186 to 98 because many of the denominations (which were extremely small [approximately 1,000 members or less] and widely dispersed across cities) were often combined with other closely affiliated denominations and then entered into the diversity index. For example, the census included 10 denominations in the denominational family Evangelistic Associations. Yet, the largest denomination had only 403 members and all 10 denominations had a total of only 1,699 members in the cities. The membership for all 10 denominations, classified as Evangelistic Associations, was totaled and entered into the diversity index for each city. In a similar fashion, the 22 Lutheran denominations were reduced to 16 denominational categories, the four Mennonite denominations were combined into one denominational category (total 1,176), the three Theosophical Societies were combined into one denominational category (total 2,065), and so forth.

Table 3. Standardized and Unstandardized Regression Coefficients for Religious Diversity, Percent Catholic, and Population Growth, with Rate of Church Adherents as the Dependent Variable

	Rate of Church Adherents			
	Equation 1		Equation 2	
	B	b	B	b
Religious diversity	.84	.76**	.79	.72**
% Catholic	1.40	1.23**	1.35	1.19**
Population growth			−.12	−.04*
$R^2 =$	.60		.62	

* At least twice its standard error.
** At least three times its standard error.

Equation 1 of Table 3 shows that both religious diversity and the percent of Catholics in the population are strong predictors of religious participation.[5] These two variables explain 60 percent of the variance and the beta for each variable is strongly positive and highly significant. As expected, the presence of Catholics and the diversity of the religious market both increase the rate of adherents in a given city.

Yet, current research suggests that religious adherence in a community is also influenced by the social context of the religious market. For example, Welch (1983) has shown that in current SMSAs, migration is an important determinant of church membership and helps to explain regional differences in the rate of church membership. This finding is consistent with numerous studies that have found migration to have negative effects on all forms of religious participation (Wuthnow and Christiano 1979; Welch and Baltzell 1983; Finke 1987). Hence, in equation 2 of Table 3 we added a measure of migration. Although we were unable to attain a

---

[5] Despite the high correlation between the independent variables, religious diversity and percent Catholic, the question of multicollinearity does not arise. Muilticollinearity results when two independent variables hold a nearly fixed relationship and lack independent variation with the dependent variable. This lack of independent, or unique, variation then reduces the precision of the coefficients by increasing the standard error, producing insignificant coefficients (Kennedy 1979; Rao and Miller 1971). Multicollinearity obscures relationships; it does not create false, strong relationships. In this case, despite high correlations among the variables, the coefficients for religious diversity and percent Catholic are highly significant. In fact, the unstandardized coefficient for religious diversity is more than six times its standard error and for percent Catholic, the unstandardized coefficient is more than nine times its standard error. While the simple correlation between religious diversity and percent Catholic is high, the pair have sufficient independent variation to allow the OLS procedure to calculate precise coefficients.

measure of residential stability, we were able to compute the rate of population growth from 1890 to 1906. Clearly, the rate of population growth adds little to the explained variance and does not alter the coefficients for percent Catholic or religious diversity. The beta is a negative, as would be expected, and significant at the .05 level, but the strength of the beta for population growth is relatively weak. When compared to the percent Catholic and the religious diversity of a city, the rate of population growth has only a modest effect on the rate of adherents. The effects of religious diversity and the percent Catholic dominate the equation.

But how does religious diversity increase the rate of adherents? If we were studying the mid-19th century, revival meetings would be the most apparent answer, for Protestants and Catholics alike (Dolan 1978; McLoughlin 1978). But after the Civil War, revivals were not the only evangelical tool used for recruitment. As Winthrop Hudson explains, "the other great evangelistic enterprise of the Protestant churches . . . was the Sunday school movement . . . within a very brief time, the Sunday school—benefiting from its surge of popularity—had begun to replace revivalism as the primary recruiting device of the churches (1981, pp. 236–37)." By 1906, the Sunday school movement was not merely an educational program for small children, but a movement designed to recruit new adult members and renew the commitment of the current membership.

How was this movement influenced by the religious pluralism of the city? Table 4 reveals the powerful effects of religious diversity on the Sunday school movement. As the pluralism of the religious market increases, the rate of Sunday school activity also increases. And, even though the Catholic church did not become active in the Sunday school movement—as can be seen by its negative correlation with the Sunday school rate—the percent of Catholics still has a positive effect on the rate of Sunday schools when we control

Table 4. Standardized and Unstandardized Regression Coefficients for Religious Diversity, Percent Catholic, and Population Growth with the Rate of Sunday Schools as the Dependent Variable

	Rate of Sunday Schools	
	B	b
Religious diversity	.98	2.50**
% Catholics	.41	.99**
Population growth	.03	.03
R^2	.43	

* At least twice its standard error.
** At least three times its standard error.

Table 5. Standardized and Unstandardized Regression Coefficients for Religious Diversity, Percent Catholic, Population Growth, and the Rate of Sunday Schools, with Rate of Church Adherents as the Dependent Variable (Includes Indirect Effects)

	Rate of Church Adherents		
	Direct Effects		Indirect Effects Through Sunday Schools
	B	b	
Religious diversity	.29	.26*	.50
% Catholic	1.14	1.01**	.21
Population growth	−.14	−.04**	.02
Sunday school.	.51	·.18**	
$R^2 =$	.76		

* At least twice its standard error.
** At least three times its standard error.

for religious diversity.[6] Hence, while Catholics did not develop a strong Sunday school program, their presence stimulated the growth of Protestant Sunday schools. Even though the effect of Catholics was much less than that of religious diversity, the coefficient was still highly significant. Once again, this suggests a powerful Protestant reaction to their perceptions of the Catholic threat. When the competition provided by religious pluralism and the threat of Catholics increased, the evangelical efforts of the Protestants—via their Sunday schools—also increased.

When the rate of Sunday schools is combined with the other variables previously used to explain church participation, the results are reassuring (Table 5). As expected, the coefficient for the rate of Sunday schools is strong, positive, and highly significant. Also, as expected, the coefficient for Catholics remains strongly positive, and the coefficient for population growth is still weakly negative. Yet the coefficient for religious diversity is still strong and significant even when the rate of Sunday schools is entered into the equation. Moreover, both religious diversity and percent Catholic have strong indirect effects on the rate of adherents. Therefore, religious diversity and percent Catholic have indirect effects on the rate of adherents by increasing the rate of Sunday schools, as well as a direct effect on religious participation.

The path diagram in Figure 1 presents both the direct and indirect effects of religious

---

[6] In fact, the Catholic equivalent of the Sunday School, the Confraternity of Christian Doctrine (CCD), did not become active in the United States until two decades later. While the first established unit of the CCD began in 1902 in New York City, few parish and diocesan CCD units began prior to the 1920s (Bryce 1986).

diversity and Catholics. As the level of diversity and percent of Catholics increase, the rate of adherents also increases. Yet the effects are not only direct ones; increases in each of the variables also leads to a sharp increase in the evangelical efforts of Protestant Sunday schools. In turn, the strength of the cities' Protestant Sunday schools has a strong effect on the level of religious participation. Thus, religious diversity not only increases religious participation by appealing to a broader segment of the market, it also increases competition and forces churches to develop effective membership recruitment and retention techniques, such as the Sunday school. Likewise, the Catholics were not only effective at recruiting the new immigrants into their churches and retaining their membership, they also threatened the Protestant status quo. This threat stimulated the growth of the Protestant churches in urban America. Regardless of their direct effects or indirect effects, both religious diversity and the percent Catholic contributed to the high level of religious involvement in urban America at the turn of the century.

DISCUSSION

Our results strongly suggest that the received wisdom about the effects of urbanism and pluralism on religion may be wrong. If so, the question arises as to how these views became enshrined as self-evident truths? We suggest this occurred because the myth of pastoral piety served the interests of both Catholic and radical European intellectuals. For a long time, Catholic writers presented the Reformation as a terrible catastrophe that shattered the moral integration of Western Civilization, eventually giving rise to such subsequent catastrophes as the French Revolution and the rise of Marxism. Implicit is the assumption that a universal church is not only theologically preferable, but also socially necessary. Explicit is the claim that faith has crumbled since the golden days of the High Renaissance when all Europeans walked secure in faith and grace. Similarly, the notion of decline fits with radical claims that the rise of science and rationalism is breaking the mystical spell that has held humanity in thrall. As the
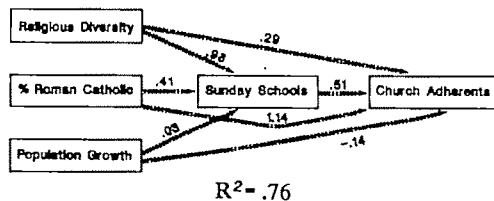
$R^2 = .76$
Fig. 1. Path Diagram of the Model of the 1906 Religious Economy

modern world unfolds, radical theorists propose, religion soon will disappear. These claims are enhanced if it can be shown that a lot of disappearing has already occurred; to this end it is much easier to overstate the piety of the past than minimize the piety of the present.

From the start, the social sciences have taken for granted the decline and eventual disappearance of religion. Perhaps no single social science thesis has come as close to universal acceptance as the belief that modernism dooms faith. The distinguished Anthony F.C. Wallace undoubtedly spoke for most of his colleagues when he confidently wrote in his celebrated book on the anthropology of religion that ". . . the evolutionary future of religion is extinction . . . belief in supernatural powers is doomed to die out, all over the world as a result of the increasing adequacy and diffusion of scientific knowledge . . . the process is inevitable" (1966, p. 265).

This dire prediction seemed almost inevitable as urban centers grew and religious pluralism mushroomed. Yet, our results indicate that both urbanism and pluralism contributed to an increase in religious mobilization.

Our results are not the first to dispute the secularization thesis. An unbiased examination of the empirical findings reported during the past decade does not support claims about the great decline in piety (Bell 1980; Stark 1981; Bahr 1982; Perkins 1984; Robertson and Chirico 1985; Hammond 1985; Stark and Bainbridge 1985; Hadden 1987). Simply because piety varies by time and place does not mean that the past was more pious than the present. Mass attendance probably was not high in medieval times. Illegitimacy was rife in communities said to be unanimously Catholic and pious (Wrigley 1969). Puritan Boston in the time of Cotton Mather had a far lower rate of church membership and attendance than it does today — far lower even than modern Las Vegas or San Francisco. And, as far as we can tell from our data on American church membership in various eras, at no time was rural America as religious as urban America (Finke and Stark 1986; Stark and Finke forthcoming).

Undoubtedly, some scholars may claim that attendance and membership are not proof of piety, and that many who didn't belong to or attend a church believed. We agree. But this may be as true today as it was in the past. Polls show that a great majority of people living in the Pacific region of the United States say they pray and claim to believe basic Christian tenets. But only about a third of the population belongs to a church, compared to more than 60 percent in most other states (Stark and Bainbridge 1985). In addition, the lower membership rates in the West are not the result of a decline. Membership

there, like membership in the rest of the nation, has risen steadily throughout the century (Finke and Stark 1986).

Bahr (1982) reports that in 1931 there was one church for every 763 residents of Muncie, Indiana (sociology's famous Middletown). By 1970 there was one church for every 473 residents — a pattern of growth that applies across the nation. This growth was not simply the result of established denominations founding new congregations, in most communities there are many more denominations than there used to be. Moreover, all through the decades when no one challenged the secularization thesis, an unbiased observer would have noticed that, somehow, the more conservative and evangelical denominations were taking over the market and causing religious participation rates to rise. For example, between 1776 and 1850 the Congregationalists saw their market share decline from nearly 21 percent to only 4, while the Baptist share rose from 17 to 21 percent, and the Methodists' (still a conservative sect movement in those days) soared from 2.5 to almost 35 percent (Stark and Finke forthcoming). The same trend of rapid evangelical growth has continued throughout this century and recently has produced a literature devoted to explaining why the conservative churches are growing as the liberal churches decline (Kelley 1972; Bibby 1978; Hoge and Roozen 1979; Warner 1983).

SUMMARY

Contrary to the pleas of the clergy and the pronouncements of social scientists, the city is surprisingly sacred and pluralism is friend, not foe, to religious mobilization. Using census data on the religious life of cities in 1906, we found them to have a higher rate of religious adherence than did the countryside. We have argued that the city not only offers easy access to churches, it also offers a variety of churches, all competing for adherents. Some sociologists have suggested that the competition of an open religious economy will undermine all forms of religious commitment, but we have argued that this competition has facilitated religious mobilization. The results support our argument. Both religious diversity and the presence of Catholics increased the rate of adherence in a city. Not only did each factor have a direct effect on the rate of adherence, they also had indirect effects by increasing the evangelical efforts of Protestant Sunday schools. Thus, a natural consequence of an open religious economy is a religious pluralism that forces each religious body to appeal successfully to some segment of the religious market, or to slide into oblivion.

Appendix. Correlations, Means, and Standard Deviations for Variables used in Regression Equations

	Rate of Adherents	Rel. Diversity	% Catholic	Pop. Growth	Sunday[a] Schools
Religious Diversity	−.40**				
% Catholic	.66**	−.88**			
Pop. growth	−.25**	−.03	−.08		
Sunday Sch.	.16*	.63**	−.46**	−.02	
Mean	.59	.67	.30	.66	1.09
Stand dev.	.15	.17	.17	.54	.43
N = 150					

* Significance level .05.
** Significance level .01.
[a] The rate of Sunday schools per person times 1,000.

REFERENCES

Bahr, Howard M. 1982. "Shifts in Denominational Demography." *Journal for the Scientific Study of Religion* 21:99–114.

Bell, Daniel. 1980. *The Winding Passage*. Cambridge, MA: ABT.

Berger, Peter. 1979. *The Heretical Imperative: Contemporary Possibilities of Religious Affirmation*. New York: Doubleday.

_____. 1967. *The Sacred Canopy*. New York: Doubleday.

Bibby, Reginald W. 1978. "Why Conservative Churches Really Are Growing." *Journal for the Scientific Study of Religion* 2:129–37.

Bryce, Mary Charles. 1986. "The Confraternity of Christian Doctrine in the United States." Pp. 27–43 in *Renewing the Sunday School and CCD*, edited by D. Campbell Wyckoff. Birmingham, AL: Religious Education Press.

Cox, Harvey. 1965. *The Secular City*. New York: Macmillan.

Dolan, Jay P. 1978. *Catholic Revivalism: The American Experience, 1830–1900*. Notre Dame: University of Notre Dame Press.

Durkheim, Émile. [1897] 1951. *Suicide*. Glencoe, IL: Free Press.

Finke, Roger. 1987. "Demographics of Religious Participation: An Ecological Approach, 1850–1971." Paper presented at the 51st Annual Meeting of the Midwest Sociological Society, Chicago.

Finke, Roger and Rodney Stark. 1986. "Turning Pews into People: Estimating 19th Century Church Membership." *Journal for the Scientific Study of Religion* 25:180–92.

_____. Forthcoming. "How the Upstart Sects Won America: 1776–1850."

Greenberg, Joseph H. 1956. "The Measurement of Linguistic Diversity." *Language* 32:109–15.

Hadden, Jeffrey K. 1987. "Toward Desacralizing Secularization Theory." *Social Forces* 65:587–611.

Hammond, Phillip E., ed. 1985. *The Sacred in a Secular Age*. Berkeley: University of California Press.

Hoge, Dean R. and David A. Roozen, eds. 1979. *Understanding Church Growth*. New York: Pilgrim Press.

Hudson, Winthrop S. 1981. *Religion in America*. New York: Scribner's.

Johnson, Douglas W., Paul R. Picard, and Bernard Quinn. 1974. *Churches and Church Membership in the United States: An Enumeration by Region, State, and County*. Washington, DC: Glenmary Research Center.

Johnson, Paul. 1976. *A History of Christianity*. New York: Harper and Row.

Kelley, Dean. 1972. *Why Conservative Churches are Growing*. New York: Harper and Row.

Kennedy, Peter. 1979. *A Guide to Econometrics*. Cambridge, MA: MIT Press.

Lieberson, Stanley. 1964. "An Extension of Greenberg's Linguistic Diversity Measures." *Language* 40:526–31.

McLoughlin, William G. 1978. *Revivals, Awakenings, and Reform*. Chicago: University of Chicago Press.

Perkins, H. Wesley. 1984. "Religious Content in American, British, and Canadian Popular Publications From 1937 to 1979." *Sociological Analysis* 45:159–65.

Rao, Potluri and Roger Leroy Miller. 1971. *Applied Econometrics*. Belmont, CA: Wadsworth.

Robertson, Roland and JoAnn Chirico. 1985. "Humanity, Globalization, and Worldwide Religious Resurgence: A Theoretical Exploration." *Sociological Analysis* 46:219–42.

Sjoberg, Gideon. 1960. *The Preindustrial City*. New York: Free Press.

Spalding, John Lancaster. [1880] 1967. "The Country and the City, 1880." in *The Church and the City, 1865–1910*, edited by Robert D. Cross. New York: Bobbs-Merrill.

Stark, Rodney. 1981. "Must All Religions Be Supernatural?" Pp. 159–77 in *The Social Impact of New Religious Movements*, edited by Bryan Wilson. New York: Rose of Sharon Press.

_____. 1985. "Church and Sect." Pp. 139–49 in *The Sacred in a Secular Age*, edited by Phillip E. Hammond. Berkeley: University of California Press.

Stark, Rodney and William Sims Bainbridge. 1985. *The Future of Religion: Secularization, Revival, and Cult Formation*. Berkeley: University of California Press.

_____. 1987. *A Theory of Religion*. New York and Bern: Peter Lang.

Stark, Rodney and Roger Finke. Forthcoming. "American Religion in 1776: A Statistical Portrait." *Sociological Analysis*.

U.S. Census Office. 1910. *Religious Bodies: 1906*, Part 1 and Part 2. Washington, DC: Government Printing Office.

Wallace, Anthony F.C. 1966. *Religion: An Anthropological View*. New York: Random House.

Warner, R. Stephen. 1983. "Visits to a Growing Evangelical and a Declining Liberal Church in 1978." *Sociological Analysis* 44:243–54.

Welch, Kevin. 1983. "Community Development and

Metropolitan Religious Commitment: A Test of Two Competing Models." *Journal for the Scientific Study of Religion* 22:167–81.

Welch, Michael R. and John Baltzell. 1984. "Geographic Mobility, Social Integration, and Church Attendance." *Journal for the Scientific Study of Religion* 23:75–91.

Wrigley, E.A. 1969. *Population and History*. New York: McGraw-Hill.

Wuthnow, Robert and Kevin Christiano. 1979. "The Effects of Residential Mobility on Church Attendance." Pp. 257–74 in *The Religious Dimension: New Directions in Quantitative Research*, edited by Robert Wuthnow. New York: Academic Press.

DEMOCRACY, ECONOMIC DEVELOPMENT, AND
INCOME INEQUALITY*

EDWARD N. MULLER

University of Arizona

Most cross-national tests of inverse association between democracy and inequality have operationalized the concept of democracy by measuring its level at a single point in time. More compelling theoretical arguments can be made for causal relationships that operate over time between (1) a country's years of democratic experience and income inequality; and (2) income inequality and the stability of democracy. Continuous and qualitative measures of years of democratic experience are estimated to have a significant negative impact on income inequality independent of economic development and other control variables for a sample of 55 countries. A very strong inverse correlation is also observed between income inequality and regime stability for a sample of 33 democracies. This association holds independent of economic development, which is found to have no direct effect on democratic stability after controlling for income inequality.

Since the early 1970s, when data on the distribution of personal income within nations at differing levels of economic development first became available, the hypothesis of an inverse relationship between political democracy and income inequality has been subjected to many cross-national tests. Most studies have focused on whether a country's level of political democracy has a negative effect on the extent of income concentration in the hands of the few, controlling for level of development. Results have been inconsistent. Rubinson and Quinlan (1977), Stack (1978), and Weede (1982) report support for the hypothesis of a negative effect of level of democracy on income inequality. Weede and Tiefenbach (1981a) report both support and lack of support for it across different samples. Bollen and Grandjean (1981) find no support for it.

Of course, if democracy and inequality are related, the causal influence could run in the other direction: inequality could have a negative effect on democracy. Rubinson and Quinlan (1977) and Bollen and Jackman (1985) have estimated models of reciprocal causation in which level of democracy and income inequality are stipulated as both cause and effect of each other. In the context of these simultaneous equation models, which include a control for economic development, Rubinson and Quinlan find evidence only of a negative effect of inequality on level of democracy. Bollen and Jackman find no significant relationship in either direction, however.

Much of this research suffers from methodological problems (e.g., sample composition, re-

liability and validity of measurement, model specification), which have been reviewed by Bollen and Jackman (1985). Greater weight should probably be given to Bollen and Jackman's work because they used a larger sample (60 cases) than others. Their indicator of level of democracy is that developed by Bollen (1980), which is superior to those used by Rubinson and Quinlan, Stack, and Weede and Tiefenbach. Bollen and Jackman take into account the possibility of reciprocal causation. Thus, the most reliable current evidence indicates that level of democracy has no systematic effect on income distribution within countries and that income inequality has no systematic effect on the level of democracy within countries.

Previous research, regardless of its methodological strengths and weaknesses, may have been barking up the wrong tree, since any measure of level of democracy at a single point in time, no matter how reliable and valid, is insensitive to a country's experience of democracy over time. Yet it is precisely longitudinal variation in democracy, measured either by the number of years that democratic institutions have existed or by the stability of democracy over a given interval, that is most likely to be associated inversely, as cause or effect, with income inequality.

Research that focuses on level of political democracy implicitly assumes that democracy can have a more or less immediate effect on inequality. All countries with a high level of democracy are expected to have a relatively low level of income inequality, regardless of the length of time that democratic institutions have existed. But if the egalitarian influence of democracy is in reality a long-term incremental effect, then relatively new democracies should not be expected to be as egalitarian as older ones, even if they have the same level of

democracy in a given year. Thus, a reason for the failure to find a significant negative effect of level of democracy on income inequality could be the confounding influence of new, inegalitarian democracies. In such countries, sufficient time has not elapsed for the institutions of democracy, operating through mediating variables such as strong trade unions and socialist parties, to have exerted an egalitarian effect on the distribution of income. *Ceteris paribus,* the confounding influence of new but inegalitarian democracies should be a function of sample size. This problem might explain in part why the relatively large-sample tests of Bollen and Grandjean (1981), Bollen and Jackman (1985), and Weede and Tiefenbach (1981a), where the cases range from' 43 to 60, find no significant negative effect of level of democracy on income inequality, while the relatively small-sample tests of Rubinson and Quinlan (1977), Stack (1978), Weede (1982), and Weede and Tiefenbach (1981a), where the cases range from 20 to 36, do report such an effect.[1]

If one accepts the argument that a relatively immediate causal effect of democracy on inequality is implausible, then one must reject the use of measures of level of democracy in research on the relationship between democracy and inequality. This has been recognized by Hewitt (1977, pp. 456–57), who proposed an index of a country's years of democratic experience as the more appropriate way of operationalizing the democracy concept. For an idiosyncratically defined sample of 25 "industrialized" nations (all of which except South Africa had experienced at least some years of democratic government),[2] Hewitt found a relatively strong negative bivariate correlation between years of democracy and the size of the share of personal income received by the richest quintile ($r = -.57$). The effect of years of democracy on income inequality was not statistically significant, however, after controlling for level of economic development.

To properly test for the presence of a negative

effect of years of democratic experience on income inequality, independent of level of economic development, one should not restrict the sample to democracies and/or to economically developed countries only. This paper performs a global test of the years of democratic experience hypothesis and also evaluates the relationship between years of democracy and income inequality in comparison with that between level of democracy and income inequality. Level of economic development must be controlled for, since there are well-established relationships between it and both income inequality (an inverted U-curve — see Ahluwalia 1974, 1976; Bornschier and Chase-Dunn 1985; Kuznets 1955, 1963; Lydall 1979; Paukert 1973; Weede 1980; Weede and Tiefenbach 1981b; World Bank 1980) and level of democracy (a positive monotonic relationship — see Adelman and Morris 1967; Bilson 1982; Bollen 1979; Coleman 1960; Dahl 1971; Jackman 1973; Olsen 1968; Russett 1965). Moreover, relatively strong association also should obtain between level of economic development and years of democracy, since almost all old democracies have very high levels of economic development.

Turning to the other side of the coin, research on the question of a negative effect of income inequality on democracy has failed to recognize that two different hypotheses can be formulated. One involves the question of *genesis:* Is a relatively egalitarian distribution of income a precondition for the inauguration of political democracy? The other involves the question of *stability:* Given that a democratic political system has been established, does a relatively egalitarian distribution of income increase the likelihood of maintaining regime stability over time?

Rubinson and Quinlan (1977) focused on the genesis hypothesis. They reasoned that a relatively egalitarian income distribution reflects the rise of a strong and autonomous bourgeoisie, which strives to establish parliamentary democracy in order to convert its economic power into commensurate political power (also see Moore 1966). Therefore, Rubinson and Quinlan postulated that a country's level of democracy will vary as an inverse monotonic function of income inequality, since countries with low income inequality, indicative of a strong middle class, will be more likely to establish democracy than those with high income inequality, where the middle class is presumably weak.

Proper testing of this version of the argument that inequality will have a negative effect on democracy requires measuring income inequality before or at least contemporaneously with the inauguration of democracy. Moreover, since countries differ greatly in the timing of the inauguration of democracy, a measure of the level of democracy that exists across all nations in the same year is an inappropriate indicator of

---

[1] Weede and Tiefenbach (1981a) conducted a relatively small sample test using data from Paukert (1973), where the number of cases was 34. They conducted relatively large sample tests using data from Ahluwalia (1974) and Bornschier and Heintz (1979), where the number of cases was 43 and 46, respectively. The study by Rubinson and Quinlan (1977) reported a significant negative. effect of level of democracy on income inequality only for a recursive model.

[2] Democracy is defined by the presence of an elected executive or an executive responsible to an elected assembly, universal manhood suffrage, and a secret ballot. Years of democratic experience is measured by the number of years during which these conditions have been fulfilled, less any period of authoritarian rule or foreign military occupation.

the dependent variable in the genesis hypothesis. Thus, tests of the genesis hypothesis that use inequality data for a relatively narrow slice of time and level of democracy measured for a single year are beside the point. The tests performed in the context of nonrecursive models by Rubinson and Quinlan (1977) and Bollen and Jackman (1985) are subject to this criticism.

The second version of the argument that inequality will have a negative effect on democracy concerns the relationship between income inequality and the stability of democracy. The stability hypothesis cannot be tested by analysis of the relationship between income inequality and the level of democracy in a country at a single point in time. The key to effective testing of it is a measure of democratic stability that is not confounded with level of democracy. Jackman (1974, p. 43) and Bollen (1980, pp. 382–84) have correctly pointed out that the index of democratic political development constructed by Cutright (1963) is an inappropriate measure of level of democracy because it confounds variation in stability with variation in level. For the same reason, it is an inappropriate measure of regime stability. The measure of democratic stability to be used here is a binary variable, according to which all countries that were democracies in 1960 are classified as unstable if during the next 20 years they experienced a transformation to an authoritarian regime because of a military or executive coup, fraudulent elections or the imposition of long-term restrictions on political liberty, or the outbreak of civil war. Otherwise, they are classified as stable. Thus, level of democracy is initially constant and subsequent variation in level is identical to variation in stability. The tests for a negative effect of income inequality on democratic stability will control for an important possible alternative cause, level of economic development (see Lipset 1959, 1981).

RESEARCH DESIGN

An effective research design for testing hypotheses about crossnational relationships between democracy, economic development, and income inequality must confront three methodological problems present in previous work (see Bollen and Jackman 1985, pp. 440–42): specification error, small sample size, and unreliable/invalid measurement of democracy and income inequality.

Specification Error

The most common kind of specification error in tests of the hypothesis of a negative effect of democracy on income inequality results from failure to allow for reciprocal causation between democracy and inequality variables measured at

approximately the same time. If level of democracy is replaced by years of democracy, however, the possibility of simultaneity bias is reduced because years of democracy are calculated from the date of the inauguration of democracy and measurements of income inequality are observed for an interval of time after the inaugural date of democracy. This does not eliminate the problem, since simultaneity bias could still be present if income inequality caused democracy to be inaugurated in the first place, and if income inequality remained essentially constant after the inauguration of democracy. In the case of old democracies, where the lag period between the inauguration of democracy and the observation of income inequality is more than a generation, it is reasonable to assume that some nontrivial change in income inequality has occurred. Of course, this assumption cannot be made for the relatively new democracies. Nevertheless, simultaneity bias is not a problem in regard to them if one assumes that there is little or no causal relationship between income inequality and the genesis of democracy.

The alternative hypothesis of a negative effect of income inequality on stability of democracy can be tested by measuring the two variables over the same interval. To infer causality, the inequality observations must precede or be simultaneous with the date of an instability event.

The other kind of specification error entails the form of the relationship between level of economic development and income inequality. Since it is nonlinear, the critical question of whether democracy affects inequality independent of level of economic development cannot properly be assessed if the relationship between economic development and inequality is represented as a montonic function. This problem only affected research in the 1970s, before Weede (1980) called attention to it.

Sample Size

Previous research on inequality and democracy has typically stipulated that the relevant population is all self-governing political units with market economies.[3] Data on income distribution have been reported for a majority of self-governing noncommunist states since the early 1960s, although the comprehensiveness of the coverage varies by region.[4] A reasonably large

---

[3] Countries with centrally planned economies are excluded because it is assumed that their data on income distribution are not comparable with data from countries with market economies (see Bollen and Jackman 1985, p. 442).

[4] It is good (two-thirds or more) for states located in Europe and North America, in Central and South

sample is one of at least approximately 50 cases.

Measurement of the Variables

Data on income distribution are heterogeneous in regard to the timing of the observations, the definition of the income recipient, the proportion of the population covered, and the nature of the data collection procedure. Any attempt to assemble comparable data must consider each of these issues.

Previous studies have assumed that a country's income distribution changes slowly and that measurements of income inequality are reasonably comparable within an interval of approximately 20 years (Ahluwalia 1976; Bollen and Jackman 1985; Lecaillon, Paukert, Morrisson, and Germidis 1984; Paukert 1973). This assumption is debatable. In testing the hypothesis of a positive effect of years of democracy on inequality, a conservative assumption is that measurements of income inequality are reasonably comparable within an interval of 11 years (1965–75).

Estimates of income distribution are typically made for households and for individuals, defined either as individual members of the population or as economically active persons. These different units of income recipients are not necessarily comparable. The issue of comparability has been investigated by Lecaillon et al. (1984, pp. 25–35). They concluded that, at least in LDCs, incomes of economically active persons are usually somewhat more concentrated than those of households or individual members of the population but that indexes of concentration such as the Gini coefficient are approximately the same for distributions of income by households and by individual members of the population. The number of cases on which these comparisons are based is small, however. Therefore, it is advisable to control for the nature of the income recipient unit (see Bollen and Jackman 1985).

Finally, comparable measurement of income distribution crossnationally requires estimates for the total population obtained from representative national surveys instead of from income tax statistics.

Data on income inequality compiled for the period 1965–75 are listed by region in Table 1, which shows the distribution of scores on the

two most commonly used measures: the size of the share of personal income received by the richest quintile, an indicator of concentration at the top of the distribution; and the Gini coefficient of concentration, an indicator of the extent to which the entire distribution deviates from perfect equality. The income recipient unit is, in general, the household; if data for households are not available in LDCs, however, then data on individuals are used. If more than one source and/or year is listed,[5] the income share is the average value. This sample of 55 countries encompasses approximately one-half of the 113 self-governing noncommunist countries that existed in 1965.[6]

A country's level of economic development is measured by energy consumption per capita in 1965 (from Ballmer-Cao and Scheidegger 1979).[7] This summary indicator is often used in preference to gross national product per capita because it avoids some of the comparability and exchange rate problems that occur when convert-

---

[5] The sources of the data are the World Bank (A—Ahluwalia 1976; J—Jain 1975; W—World Bank 1980–1985), the International Labour Organization (L—Lecaillon et al. 1984; P—Paukert 1973), the Organization for Economic Development and Cooperation (S—Sawyer 1976), and the United Nations Economic Commission for Latin America (C—Comisión Económica para América Latina 1983). Multiple observations were averaged if they were at least three years apart.

[6] This set of income-inequality measurements excludes data from several countries that were included in the study by Bollen and Jackman (1985). A 1966 measurement for Denmark is excluded because it is based on tax statistics, which tend to exaggerate inequality in comparison with data from population surveys (see Sawyer 1976, p. 23). A 1969 measurement for the Dominican Republic is excluded because it is from a survey carried out only in the capital city. A 1970 measurement for Ecuador is excluded because, although there is no indication of exactly how the data were obtained, they do not seem to be based on a sample survey (see Luzuriaga and Zuvekas 1983, p. 8). A circa 1970 measurement for Pakistan is excluded because the data seriously underestimate the incomes of higher income groups (see van Ginnekan 1976, p. 14). A 1970 measurement for Uganda is excluded because it refers only to African male wage-earners (a relatively small proportion of the population). Also, measurements for Liberia, Morocco, Niger, and Syria from a 1980 World Bank Memorandum have not been used because data for these countries were not reported in the first World Bank compendium of income distribution data (Jain 1975), nor do they appear in subsequent publications by the World Bank (World Development Report, annually 1980–85). Finally, data for five countries (Benin, Chad, Iraq, Lebanon, and Madagascar) are not used because they predate 1965.

[7] Data for Barbados and Gabon are from Taylor and Hudson (1972).

America, and in Asia. It is unsatisfactory (approximately one-third) for states located in the Middle East and North Africa and in Sub-Saharan Africa. If region were an important causal variable, differences in regional coverage would create a problem. Here, region is assumed to be an unimportant causal variable.

Table 1. Share of Personal Income Received by the Richest Quintile of Households/Individuals and Gini Coefficient

	Income Share of Upper Quintile, 1965–75				Income Share of Upper Quintile, 1965–75		
	(Source:Year)	%	Gini		(Source:Year)	%	Gini
Europe & N. America				*Middle East & N. Africa*			
Belgium	(W:75)	39.8	.30	Egypt	(A:65;W:74)	47.5	.39
Canada	(W:69)	41.0	.34	Iran	(L:71)	62.7	.52
France	(W:70,75)	46.4	.38	Tunisia	(A:70)*	55.0	.46
Germany, West	(S:70;W:74)	45.2	.36	Turkey	(A:68;W:73)	58.6	.49
Ireland	(W:73)	39.4	.30	*Sub-Saharan Africa*			
Italy	(W:69)	46.5	.38	Gabon	(A:68)*	67.5	.55
Netherlands	(W:67,75)	40.0	.30	Ivory Coast	(A:70)*	57.2	.49
Norway	(W:70)	37.3	.30	Kenya	(W:76)	60.4	.51
Portugal	(W:73–74)	49.1	.40	Malawi	(W:67–68)	50.6	.34
Spain	(A:65;W:74)	45.2	.36	Senegal	(L:70)	60.9	.47
Sweden	(W:72)	37.0	.29	Sierra Leone	(W:67–69)	52.5	.42
United Kingdom	(S:67;W:73)	39.2	.31	South Africa	(L:65)	58.0	.53
United States	(W:72)	42.8	.36	Sudan	(W:67–68)	49.8	.41
Central & S. America				Tanzania	(W:69)	50.4	.39
Argentina	(W:70)	50.3	.41	Zimbabwe	(L:69)	68.0	.57
Barbados	(J:69–70)*	44.0	.34	*Asia*			
Bolivia	(P:68)	61.0	.49	Australia	(W:66–67,75–76)	43.0	.34
Brazil	(W:72)	66.6	.56	India	(W:64–65,75–76)	49.2	.38
Chile	(W:68)	51.4	.43	Indonesia	(J:71)*	52.0	.39
Colombia	(L:74)	58.5	.48	Japan	(W:69)	41.0	.30
Costa Rica	(A:71)	50.6	.41	Korea, South	(J:66;L:70)	42.7	.33
El Salvador	(J:65–67,69)*	54.4	.46	Malaysia	(W:70,73)	56.4	.47
Guatemala	(C:70)*	58.8	.46	New Zealand	(S:65–66)	41.4	.31
Honduras	(W:67)	67.8	.57	Philippines	(W:70–71)	54.0	.43
Mexico	(A:69)	64.0	.52	Sri Lanka	(W:70;L:73)	44.5	.34
Nicaragua	(C:70)*	60.0	.51	Taiwan	(W:71)	39.2	.28
Panama	(W:70)	61.8	.54	Thailand	(L:69;W:75–76)	53.4	.44
Peru	(W:72)	61.0	.54				
Trinidad & Tobago	(W:75–76)	50.0	.42				
Uruguay	(A:67)	47.4	.40				
Venezuela	(W:70)	54.0	.47				

* Income recipient unit is the individual.

ing GNP to a common currency unit such as the U.S. dollar.

Also, following Bollen and Jackman (1985), I control for the proportion of the population that is young, as measured by the percentage 0–14 years of age (World Bank, vol. 2, 1983), and for world-system position, as measured by the classification of Snyder and Kick (1979) with revisions by Bollen (1983). Income inequality is expected to be related positively to youth of population and a semiperipheral or peripheral world-system position.

The most reliable and valid indicator of a country's level of democracy is Bollen's (1980) Political Democracy Index (PDI). Scores on the PDI range from 0–100 (low to high). The first column of Table 2 lists the 1965 PDI value for countries with information on income inequality during 1965–75.

The years of democratic experience for each of these countries can be calculated by determining the date when democracy was established and any years since when demo-

cratic rule was either interrupted temporarily or replaced by a nondemocratic regime. For a functioning democratic regime to be established, (1) the executive must be elected or be responsible to an elected assembly in (2) at least two consecutive free and fair competitive elections in which (3) at least approximately a majority of the adult population has the right to vote, and during which (4) the rights of freedom of speech and assembly are respected.[8] If these criteria are fulfilled, then the date of the first election is the inaugural year of democracy, as listed in the second column of

---

[8] The first, third, and fourth criteria are standard components of contemporary definitions of democracy (cf. Bollen 1980, pp. 371–73; Dahl 1971, pp. 1–9; Hewitt 1977, pp. 456–57; Powell 1982, p. 3). The stipulation of at least two consecutive free and fair elections is included to eliminate transitory "experiments" in democracy whose duration is too brief to be considered a functioning regime.

Table 2. Bollen Political Democracy Index and Years of Democracy to Most Recent Date of Income Distribution Measurement

	Political Democracy Index, 1965	Date of Inauguration of Democracy	Interruptions/ Breakdowns of Democracy	Years of Democracy to:
Europe & N. America				
Belgium	100	1919	(1940–44)	1975: 51
Canada	100	1898		1969: 71
France	91	1875	(1940–44;1958)	1975: 94
Germany, West	89	1949		1974: 25
Ireland	97	1923	.	1973: 50
Italy	97	1946		1969: 23
Netherlands	100	1918	(1940–44)	1975: 52
Norway	100	1898	(1940–44)	1970: 67
Portugal	39	—		1974: 0
Spain	10	—		1974: 0
Sweden	100	1917		1972: 55
United Kingdom	99	1918		1973: 55
United States	92	1870		1972:102
Central & S. America				
Argentina	53	1916	(1930→*)	1970: 0[a]
Barbados	100	1961		1970: 9
Bolivia	36	—		1968: 0
Brazil	61	1945	(1955;1964→*)	1972: 17
Chile	97	1949	(1973→*)	1968: 19
Colombia	71	1974		1974: 0
Costa Rica	90	1949		1971: 22
El Salvador	72	—		1969: 0
Guatemala	40	1945	(1954→*)	1970: 8
Honduras	50	—		1967: 0
Mexico	75	—		1969: 0
Nicaragua	55	—		1970: 0
Panama	77	1960	(1968→*)	1970: 7
Peru	87	1956	(1962;1968–80*)	1972: 10
Trinidad & Tobago	85	1961		1975: 14
Uruguay	100	1919	(1933;1942;1973→*)	1967: 46
Venezuela	73	1958		1970: 12
Middle East & N. Africa				
Egypt	39	—		1974: 0
Iran	45	—		1971: 0
Tunisia	64	—		1970: 0
Turkey	76	1950	(1960–61;1971–73;1980→*)	1973: 18
Sub-Saharan Africa				
Gabon	45	—		1968: 0
Ivory Coast	46	—		1970: 0
Kenya	58	—		1974: 0
Malawi	58	—		1968: 0
Senegal	54	—		1970: 0
Sierra Leone	75	—		1969: 0
South Africa	65	—		1965: 0
Sudan	38	—		1968: 0
Tanzania	63	—		1969: 0
Zimbabwe	—	—		1969: 0
Asia				
Australia	100	1892		1975: 83
India	91	1947	(1975–76)	1975: 27
Indonesia	10	—		1971: 0
Japan	100	1952		1969: 17
Korea, South	53	—		1970: 0
Malaysia	80	1957	(1969→*)	1973: 11
New Zealand	100	1879		1966: 87
Philippines	93	1946	(1972→*)	1971: 25
Sri Lanka	86	1948	•	1973: 25
Taiwan	23	—		1971: 0
Thailand	17	—		1975: 0

* Democracy replaced by an authoritarian regime.
[a] Reverts to zero because the last year of democracy occurred more than 20 years before the 1965–75 interval.

Table 2.[9] If the criteria for the establishment of a functioning democracy were not fulfilled at any time before 1975, no inaugural date of democracy is given.[10]

A country's years of democracy, listed in the fourth column, is calculated to the most recent date of the income-inequality measurement; it is the difference between that date and the inaugural year of democracy, less any years of interruptions or breakdowns of democracy.[11] Interruptions include the temporary suspension of democratic liberties, as well as periods when the government is controlled by a foreign power.[12] Breakdowns occur when democratic regimes are replaced internally by authoritarian rule or lose authority over a large proportion of the population due to civil war.

Countries with many years of democratic experience also tend to be politically stable countries. The years of democracy variable, therefore, will be correlated with general political system stability.[13]

---

[9] Data from Table 2 of Hewitt (1977) have been used to determine the inaugural date in most countries where democracy was established before World War II. In *Australia, France, Ireland,* and *Sweden,* the inaugural date is the year when the executive was first elected or became responsible to an elected assembly. In *Argentina, Belgium, Canada,* the *Netherlands, New Zealand, Norway,* the *United Kingdom,* and the *United States,* it is the year when universal adult manhood suffrage was instituted (giving approximately a majority of the population the right to vote). In the case of *Uruguay,* the year given coincides with the constitution that went into effect in 1919 (see Weinstein 1975, pp. 58–65). Democracy was inaugurated in Germany in 1918, but Germany ceased to exist as a state in 1945; *West Germany* is treated as a new state, which became independent in 1949. Among the other post-World War II democracies, the inaugural dates are as follows: in *Barbados,* after full internal self-government was granted by Britain; in *Brazil,* when the civilian authorities first allowed a genuinely free and fair election to be held (see Wesson and Fleischer 1983, pp. 14–21); in *Chile,* when women were granted the right to vote (literacy restrictions were still in effect); in *Colombia,* after the expiration of the National Front Agreement (see Kline 1985, pp. 256–59); in *Costa Rica,* after the revolution of 1948; in *Guatemala,* after an authoritarian civilian government was overthrown by the military, which then sponsored a transition to democracy (see Schlesinger and Kinzer 1983, pp. 25–35); in *India,* after independence; in *Italy,* after Allied control was relinquished; in *Japan,* after American control was relinquished; in *Malaysia,* after independence; in *Panama,* when the civilian authorities first allowed a genuinely free and fair election to be held (see LaFeber 1979, pp. 132–59); in *Peru,* after a military government abdicated in favor of democratic rule (see Cotler 1978, passim); in the *Philippines,* after independence; in *Sri Lanka,* after independence; in *Trinidad and Tobago,* after full internal self-government was granted by Britain; in *Turkey,* when the civilian authorities first allowed a genuinely free and fair election to be held (see *Encyclopedia Britannica* 1966, vol. 22, pp. 610–11); in *Venezuela,* after a military government abdicated in favor of democratic rule (see Wiarda 1985, pp. 306–10). The principal general source consulted for information on the inauguration of democracy in countries not covered by Hewitt is the *Encyclopedia Britannica* (1966); useful supplementary general sources include Coleman (1960), Dahl (1971), Lijphart (1977), Powell (1982), and Rustow (1967).

[10] All countries defined as under democratic rule in 1965 on the years of democracy index also have 1965 PDI scores in the upper quarter of the scale (75 or more), except for Venezuela. All of the countries defined as not having inaugurated democracy have PDI scores in the lower three-quarters of the scale (< 75), except for Mexico and Sierra Leone. Venezuela, Mexico, and Sierra Leone are borderline cases, with PDI scores of 73, 75, and 75, respectively. Although a relatively high level of violence accompanied elections held in Venezuela during the late 1950s and early 1960s, country specialists such as Levine (1978), Martz (1968), and Wiarda (1985) judge them to have been competitive and free. By contrast, although

scholars writing in the 1960s disputed the nature of the regime in Mexico (it was classified as semi-competitive by Coleman [1960] but as a democracy by Rustow [1967]), the current consensus, including the revised opinion of Rustow (1970, p. 349), is that Mexico under the PRI has a hegemonic party system (Sartori 1976, p. 234) that does not permit free and fair competitive elections (see, e.g., Stevens 1985, pp. 435–36). Sierra Leone was classified by Taylor and Hudson (1972) as having held an election in 1965 that displayed extreme deviation from the competitive and free norm. Moreover, since 1973 Sierra Leone has consistently received a rating by Gastil (1986) of only "partly free" in regard to political rights and civil liberties. El Salvador also is a borderline case, with a PDI score (72) about the same as Venezuela. In El Salvador, reasonably free and fair competitive elections for the legislature and the presidency were held in 1966 and 1967. But after 1967 the Central Council of Elections came under military influence and subsequent elections were rigged in favor of candidates preferred by the military (see McDonald 1985, pp. 537–38; Peterson 1968, p. 88). Therefore, El Salvador is not considered ever to have inaugurated a functioning democracy.

[11] In the case of Argentina, however, where democracy was inaugurated in 1916 and lasted until 1930, these years are discounted because they occurred so long before the measurement period for income inequality. Although democracy has not been effectively restored in Argentina since 1930, the current regime is democratic, and if it survives through a second free and fair election then a functioning democracy finally could be said to have been reestablished.

[12] Domestic interruptions of democracy occurred during the transition from a parliamentary to a presidential form of government in France in 1958; when the military intervened as an "arbitrator" during crises that occurred in Brazil in 1955, in Peru in 1962, and in Turkey in 1960–61 and 1971–73; when parliament was dissolved and constitutional reforms were promulgated by executive fiat and ratified by plebiscite in Uruguay in 1933 and again in 1942; and when Indira Gandhi proclaimed a state of emergency and suspended democratic liberties in India during 1975–76. Democracy was interrupted by German military occupation during World War II in Belgium, Denmark, France, Luxembourg, the Netherlands, and Norway.

[13] The years of democracy measure does not *confound* stability with democracy because stability is not part of

Political system stability can be measured by the number of years that a country has been governed under the rules of the current constitution (see Bollen 1980, p. 383), calculated as the difference between the most recent date of the income-inequality measurement and the inaugural year of the constitution (from Table 2.1 of Taylor and Hudson 1972). The correlation between years of democracy and age of constitution for the 55 countries listed in Table 2 is +.63. Since any stable political system might tend to have lower inequality, regardless of whether it is a democracy or not, system stability in general should be controlled for when testing the years of democracy hypothesis.

DEMOCRACY AND INCOME INEQUALITY

The first hypothesis predicts that a country's accumulated years of democratic experience will have a stronger negative effect on income inequality than the level of democracy that exists at a single point in time. The simple correlations between years of democracy and the two indicators of inequality, the size of the upper 20 percent income share and the Gini coefficient of concentration, are −.61 and −.55, respectively. By contrast, the respective correlations between the 1965 PDI and the indicators of inequality are −.43 and −.37. Although these bivariate results support the democratic-experience hypothesis, a plot of the inequality variables by years of democracy shows that the form of the relationship departs substantially from linearity and is best described as a reversed J-curve. Logging is one possible transformation, but it does not improve the fit. The correlations between the natural logarithm of years of democracy[14] and the inequality variables are −.56 and −.51, respectively. An alternative transformation is to classify the years of democracy variable into a set of rank-ordered categories: nondemocracies (no years of democratic experience); new democracies (less than a generation—20 years—of democratic experience): relatively new democracies (more than one but less than two generations of democratic experience): relatively old democracies (more than two but less than three generations of democratic experience); and very old democracies (three or more generations of democratic experience). The predictions about income

---

the operational definition. According to the years of democracy measure, countries with many years of democracy are not necessarily stable and those with few years of democracy are not necessarily unstable. Uruguay is an example of an unstable democracy that nevertheless has had many years of democracy; Venezuela is an example of a stable democracy that nevertheless has had few years of democracy.

[14] An increment of 1 is added because the log of zero is undefined.

come inequality based on these ordinal categories of experience of democracy are (1) little or no difference in inequality between nondemocracies and new democracies; (2) progressively larger negative effects for relatively new and relatively old democracies; and (3) a negative effect for very old democracies of similar size to that for relatively old democracies.

Regressions of income inequality on level and experience of democracy are reported in Table 3. Individual Income, scored 1 for countries with measurements of income distribution by individuals and 0 for countries with measurements of income distribution by households, is controlled for in all equations. Individual Income is never statistically significant at even the .05 level, and its t-value is always very small (less than ± 1.0).

For the upper 20 percent income share, accuracy of prediction (R^2 adjusted for sample size and the number of explanatory variables) is twice as high for the equation with years of democracy (3.2a) compared with that for level of democracy (3.1a). For the Gini coefficient of concentration, R^2 is more than twice as high for the equation with years of democracy (3.2b) than that for level of democracy (3.1b). These results strongly support the argument that the egalitarian influence of democracy is a gradual process, resulting from a country's accumulated years of democratic experience, rather than a relatively immediate effect of the level of democracy in a given year.[15]

---

[15] The implicit assumption of the years of democracy measures is that what counts for reduction of income inequality is simply the presence or absence of democratic institutions over time, not variation in the extent to which democratic institutions maximize democratic procedures. However, it could be argued that the years of democracy measures lose important information because they are insensitive to variation in level of democracy. It is obviously not desirable to incorporate variation in level of democracy directly into the years of democracy measures, since this would entail a procedure like that of Cutright (1963), which confounds level and stability—or, more precisely, years—of democracy. But the question of the relevance of level of democracy separate from years of democracy can be tested empirically by including the PDI with years of democracy in additive and interaction models. When this is done, the following results are obtained: (1) a country's level of democracy has no effect on income inequality when included additively with continuous and qualitative measures of experience of democracy; (2) the multiplicative interaction term, PDI × Years of Democracy, raises R^2 by only a trivial amount relative to that for the Years of Democracy variable alone; (3) there is no change in R^2 for the multiplicative interactions of the PDI with the qualitative categories of democratic experience relative to that for the categories of democratic experience alone. These results indicate that level of democracy is irrelevant, once experience of democracy is taken into account.

Table 3. Regressions[a] of Income Inequality on Level of Democracy and Experience of Democracy

	Upper 20% Income Share, 1965-75			
	(3.1a)	(3.2a)	(3.3a)	(3.4a)
Intercept	59.71	60.54	58.47	58.32
Individual income	+3.09 (0.95)	+0.69 (0.23)	+0.29 (0.10)	+0.32 (0.11)
Political Democracy Index, 1965	−0.13** (−3.07)			
ln years of democracy		−2.01** (−3.21)		
Very new democracy			−0.024 (−0.01)	−0.04 (−0.02)
Relatively new democracy			−5.58** (−1.76)	−5.64** (−1.79)
Relatively old democracy			−12.85** (−3.09)	
Very old democracy			−10.64** (3.06)	
Old democracy				−11.84** (−4.30)
ln age of constitution		−1.87** (−2.13)	−1.31* (−1.51)	−1.25* (−1.47)
Adj. R^2 = (N =)	.17 (54)	.34 (55)	.41 (55)	.42 (55)

	Gini Income Inequality, 1965-75			
	(3.1b)	(3.2b)	(3.3b)	(3.4b)
Intercept	0.49	0.50	0.48	0.48
Individual income	+0.031 (0.98)	+0.0089 (0.29)	+0.0036 (0.13)	+0.0040 (0.14)
Political Democracy Index, 1965	−0.0010** (−2.50)			
ln years of democracy		−0.017** (−2.69)		
Very new democracy			+0.0096 (0.39)	+0.0094 (0.39)
Relatively new democracy			−0.055** (−1.70)	−0.056** (−1.73)
Relatively old democracy			−0.12** (−3.48)	
Very old democracy			−0.09** (−2.52)	
Old democracy				−0.10** (−3.71)
ln age of constitution		−0.017** (−1.88)	−0.011* (−1.29)	−0.011 (−1.22)
Adj. R^2 = (N =)	.12 (54)	.27 (55)	.35 (55)	.35 (55)

[a] t ratio in parentheses.
* $p \leq .10$, one-tailed.
** $p \leq .05$, one-tailed.

Scatterplots of the income inequality variables by years of democracy indicate that, on the average, at least approximately 20 years of democratic experience are required for the egalitarian effect to occur; a second generation of democratic experience enhances this egalitarian effect; and countries with three or more generations of democratic experience are about as egalitarian as those with two generations of democratic experience. Equations 3.3a and 3.3b replace the logged years of democracy measure by qualitative categories of democratic experience. The intercept is the expected level of income inequality for countries with no experience of democracy after the income recipient unit and the age of the constitution are controlled for. The regression coefficients for very new democracies are nonsignificant, which confirms the expectation that countries with less than a generation of democracy are no less inegalitarian, on the average, than nondemocracies. By contrast, the regression coefficients for relatively new democracies, relatively old democracies, and very old democracies are all significant and the differences in the magnitude of the effects are as expected, with both relatively old and very old democracies having about twice the egalitarian influence of relatively new democracies. Since there is little difference between the egalitarian effects of the relatively old democracies (Belgium, Denmark, Ireland, the Netherlands, Sweden, the United Kingdom, and Uruguay) and the very old democracies (Australia, Canada, France, New Zealand, Norway, and the United States), these are combined into the category 'Old Democracy' in equations 3.4a and 3.4b, where the nonsignificant Very New Democracy category is deleted. Comparison of R^2 for 3.4a/3.3a with 3.2a and for 3.4b/3.3b with 3.2b shows that accuracy of prediction is improved when the qualitative categories of democratic experience are substituted for the continuous measure.

The simple correlations between age of constitution (logged) and the two inequality variables, the upper 20 percent income share and the Gini coefficient of concentration, are $-.49$ and $-.45$, respectively. When this indicator of political system stability in general is controlled for, the years of democracy measures are always significant at the .05 level. Age of constitution also has significant negative effects on inequality in most of the equations, but they are typically weak, as indicated by the relatively small t-values.

The equations reported in Table 4 take into account the inverted U relationship between economic development and income inequality, as well as the proportion of young people in the population and world-system position. When the correct specification of the economic develop-ment effect is controlled for, level of democracy is estimated to have no significant effect on income inequality (4.1a and 4.1b). This result supports the conclusion of Bollen and Jackman (1985). By contrast, the continuous and qualitative measures of experience of democracy are estimated to have significant negative effects on income inequality independent of level of economic development (4.2a and 4.2b; 4.4a and 4.4b). These findings contradict that of no significant effect (for years of democracy) reported by Hewitt (1977). Since these results are based on a sample of 55 cases, they are more reliable than Hewitt's, which were based on a sample of only 18 cases with data on income inequality.

Including youth of population and world-system status in the equations reduces the number of cases to 50.[16] The experience of democracy measures again show significant negative effects on income inequality independent of youth of population and world-system status (4.3a and 4.3b; 4.5a and 4.5b). The results for youth of population and world-system status are consistent with those reported by Bollen and Jackman (1985): the former has a significant positive effect on inequality, while the latter has no effect.

The Individual Income variable again is not significant when level of economic development is controlled for (and is, therefore, deleted so as not to waste a degree of freedom when the population and world-system controls are introduced). Age of Constitution is significant in the equations with the continuous years of democracy variable (4.2a and 4.3a; 4.2b and 4.3b) but not when experience of democracy is expressed by qualitative categories and the control for level of economic development is introduced (4.4a; 4.4b).

The results of the equations estimated in Table 4 show that the inverse association between years of democracy and income inequality is not spurious due to level of economic development. This is an important finding, given the relatively close association that exists between the ln Energy Consumption p.c. and ln Years of Democracy variables ($r = +.67$).

INCOME INEQUALITY
AND DEMOCRACY

The next question is whether income inequality has a negative causal effect on democracy. Previous research has addressed this question by estimating the effect of income inequality on

---

[16] Barbados, Malawi, Tanzania, and Zimbabwe are missing on the world-system classification; Taiwan is missing on population age structure.

AMERICAN SOCIOLOGICAL REVIEW

Table 4. Regressions[a] of Income Inequality on Level of Democracy, Experience of Democracy, Economic Development, Youth of Population, and World-system Status

	Upper 20% Income Share, 1965–75				
	(4.1a)	(4.2a)	(4.3a)	(4.4a)	(4.5a)
Intercept	21.28	18.13	−13.09	30.63	1.52
Individual income	+2.13 (0.75)	+0.60 (0.21)		+0.25 (0.09)	
Political Democracy Index, 1965	−0.33 (−0.75)				
ln years of democracy		−1.22** (−1.71)	−1.01** (−1.64)		
Relatively new democracy				−5.29** (−1.77)	−4.79** (−1.83)
Old democracy				−8.16** (−2.40)	−6.29** (−2.10)
ln age of constitution		−1.36* (−1.51)	−1.05* (−1.38)	−1.03 (−1.15)	
ln energy consumption p.c., 1965	+13.28** (2.15)	+14.31** (2.36)	+16.74** (2.26)	+9.67* (1.53)	+12.65** (1.64)
ln energy consumption p.c.^{2}	−1.24** (−2.59)	−1.21** (−2.57)	−1.29** (−2.24)	−0.83** (−1.64)	−1.02** (−1.68)
% population < 15 years, 1965			+0.51** (3.32)		+0.42** (2.63)
Semiperiphery			−3.12 (−0.84)		−2.62 (−0.72)
Periphery			−1.18 (−0.29)		−0.72 (−0.18)
Adj. R^2 =	.36	.41	.60	.44	.61
(N=)	(54)	(55)	(50)	(55)	(50)

	Gini Income Inequality, 1965–75				
	(4.1b)	(4.2b)	(4.3b)	(4.4b)	(4.5b)
Intercept	0.008	−0.013	−0.23	0.12	−0.088
Individual income	+0.023 (0.80)	+0.0072 (0.25)		+0.004 (0.13)	
Political Democracy Index, 1965	−0.0003 (−0.66)				
ln years of democracy		−0.013** (−1.73)	−0.011** (−1.90)		
Relatively new democracy				−0.054** (−1.78)	−0.054** (−2.11)
Old democracy				−0.084** (−2.44)	−0.67** (−2.28)
ln age of constitution		−0.015** (−1.64)	−0.014** (−1.86)	−0.012 (−1.27)	
ln energy consumption p.c., 1965	+0.16** (2.50)	+0.17** (2.71)	+0.16** (2.17)	+0.12** (1.85)	+0.11* (1.50)
ln energy consumption p.c.^{2}	−0.014** (−2.80)	−0.013** (−2.77)	−0.011** (−2.01)	−0.0093** (−1.81)	−0.0085* (−1.43)
% population < 15 years, 1965			+0.0057** (3.83)		+0.0047** (3.03)
Semiperiphery			−0.024 (−0.66)		−0.017 (−0.49)
Periphery			−0.016 (−0.41)		−0.010 (−0.25)
Adj. R^2 =	.27	.34	.59	.38	.59
(N=)	(54)	(55)	(50)	(55)	(50)

[a] t-ratio in parentheses.
* $p \leq .10$, one-tailed.
** $p \leq .05$, one-tailed.

level of democracy in simultaneous-equation models. Such nonrecursive models are inappropriate, however, because they fail to correctly represent the causal processes involved. On the one hand, income inequality may have a negative impact on level of democracy because it reduces the likelihood of the inauguration of democracy. This is the genesis hypothesis. Proper testing of it requires that income inequality be measured either before or contemporaneously with the inauguration of democracy. Countries that inaugurated democracy long before the timing of inequality measurements would have to be excluded. On the other hand, income inequality may have a negative impact on level of democracy because it reduces the likelihood of the stability of democracy. Proper testing of this stability hypothesis requires that income inequality be measured either before or contemporaneously with the breakdown of democracy. Countries that are not democracies should be excluded because the stability of democracy hypothesis refers to democracies only, and inclusion of nondemocracies in the analysis would add irrelevant and potentially confounding cases. Thus, the simultaneous-equation method cannot be applied mechanically to the question of the causal effect of inequality on democracy because that entails different hypotheses with different dependent variables and different populations of cases.

The Inauguration of Democracy

According to the genesis hypothesis, countries with a relatively egalitarian distribution of income will be more likely to inaugurate democracy than those with an inegalitarian income distribution. During 1945–61, a total of 23 countries inaugurated democracy.[17] Excluding pre-1945 democracies, information on the distribution of income during 1945–61 is available for 27 countries, 14 of which are among the set that inaugurated democracy. Since most of the income distribution measurements are circa 1960, and most of the countries that inaugurated democracy did so before 1960, most of the inequality observations were made

after the inauguration of democracy. However, if one assumes that the distribution of income did not change significantly in these countries during 1945–61, then the hypothesis of an inverse relationship between income inequality and the inauguration of democracy can be tested.

Table 5 shows the relationship between income inequality, as measured by the size of the upper-quintile share,[18] and the binary variable, presence or absence of inauguration of democracy. Each country's level of democracy circa 1960, as measured by the Political Democracy Index, is given in parentheses.[19] The point-biserial correlation between income inequality and the inauguration of democracy is approximately zero $(-.07)$; and it also is approximately zero $(-.09)$ for the relationship between income inequality and level of democracy. There is no support for the genesis hypothesis in these data.

The Stability of Democracy

A succinct statement of the stability hypothesis has been given by Dahl (1971) in the following propositions:

[1] In a society that already has a regime with public contestation, extreme inequalities increase the chances that competitive politics will be displaced by a hegemony. [1a] Polyarchies are particularly vulnerable to the effects of extreme inequalities. Extreme inequalities in the distribution of key values are unfavorable to competitive politics and to polyarchy because this state of affairs: [2] is

---

[17] In addition to the 17 countries identified in Table 2 as having inaugurated democracy during 1945–61, democracy also was established during this period in *Ecuador* in 1948 after the military relinquished power and the constitution of 1946 went into effect; in *Greece* in 1950 after the end of a civil war; in *Israel* in 1949 after independence and the armistice that ended the war with surrounding Arab states; in *Jamaica* in 1953 after full internal self-government was granted by Britain; in *Lebanon* in 1945 after French military control was relinquished; and in *Malta* in 1947 after full internal self-government was granted by Britain.

[18] The data sources are Ahluwalia (1976), Jain (1975), and Kuznets (1963).

[19] All countries defined as having inaugurated democracy have PDI scores in the upper quarter of the scale, except for Guatemala, which experienced a breakdown of democracy in 1954. All but three of the countries defined as not having inaugurated democracy have PDI scores in the lower three-quarters of the scale (the exceptions are Madagascar, Mexico, and Tunisia). See note 10 for the reason why Mexico is not considered to have inaugurated democracy. Madagascar's extremely high score on the PDI (92!) is quite puzzling. On the Electoral Irregularity and Press Freedom variables (from Taylor and Hudson 1972), which are components of the PDI, Madagascar received an Electoral Irregularity rating of "D," defined as insufficient information available for a reliable judgment about electoral freedom, and Madagascar also was not even rated on the Press Freedom variable. Thus, Madagascar's PDI score could hardly be considered reliable. Also, since Tunisia was rated on the Electoral Irregularity variable as having held an election in 1964 that displayed significant deviation from the free and fair norm, and since Tunisia's Press Freedom score was $-.7$ (the global median is $+1.0$), it seems that Tunisia's PDI of 84 is excessively high.

Table 5. Inauguration of Democracy by Income Inequality

Upper 20%-Income Share, 1945–61		Inauguration of Democracy, 1945–61[*]	
(Source:Year)	%	No	Yes
(J:60)	70.8	Gabon(63)	
(J:61)	69.2		Peru(83)
(A:56)	68.0	Iraq(16)	
(J:60)	62.5	Senegal(49)	
(J:60)	62.1		Brazil(91)
(K:57)	61.7	Mexico(80)	
(J:58)	61.5		Jamaica(91)
(A:55–60)	61.0		Lebanon(84)
(J:60)	60.1	Madagascar(92)	
(J:61)	58.6		Costa Rica(91)
(K:53)	56.4	Colombia(70)	
(K:47–48)	55.4		Guatemala(70)
(J:60)	55.3		Panama(75)
(J:56)	54.8		Philippines(93)
(J:61)	54.8	Tunisia(84)	
(J:53)	53.7		Sri Lanka(94)
(K:46)	52.1	El Salvador(54)	
(J:59)	51.8	Ivory Coast(43)	
(J:59)	51.7	Benin(55)	
(J:60)	51.7		India(94)
(K:51–52)	51.6		Barbados(100)[a]
(J:61)	50.9	Argentina(63)	
(J:59–60)	50.9	Taiwan(23)[a]	
(J:57–58)	49.6		Malaysia(84)
(K:48)	48.1		Italy(97)
(K:50)	48.0		West Germany(88)
(J:61)	44.8	Chad(58)	
	[r = −.07]		

[*] 1960 PDI score in parentheses.
[a] 1965 PDI score; 1960 missing.

equivalent to extreme inequality in the distribution of key political resources and is likely to generate resentments and frustrations which weaken allegiance to the regime. (p. 103)

A weakening of support for a democratic regime due to the presence of high levels of economic inequality is equivalent to what Nordlinger (1977, p. 93) calls "legitimacy deflation." Students of the causes of coups d'etat (e.g., Finer 1962; Nordlinger 1977; Welch and Smith 1974) attach great importance to the legitimacy variable. Given the presence of interventionist motives, a low level of legitimacy is considered a condition that causes such motives to become coup attempts. By contrast, a high level of legitimacy is thought to operate as a psychological barrier that inhibits the military or the executive from usurping power. As rational actors, they recognize that a coup attempt against a legitimate regime is likely to generate opposition, not only from important segments of the political and social elite, but also from within the officer corps itself, which reduces the likelihood of success. Moreover, even if a coup were successful, the overthrow of a legitimate regime would risk provoking mass protests,

armed resistance, and civil war, all of which are contrary to the strongly held military preference for domestic order. In short, if it can be assumed that the legitimacy of democratic government is eroded by a high level of income inequality, then it follows that a democratic regime in any extremely inegalitarian society is highly susceptible to overthrow by coup d'etat.

A total of 40 countries were under democratic rule in 1961. These comprise the relevant population of cases for testing the stability hypothesis for the period 1960–80. Information on income inequality during 1960–80 is available for 33 of these democracies, a sample consisting of 82.5 percent of the population.[20]

---

[20] See Table 6 for a listing of these countries. The cases missing data on income inequality are Austria, Ecuador, Greece, Iceland, Jamaica, Luxembourg, and Malta. Of the 40 countries classified as democracies in 1961, 38 score in the upper quarter of the Bollen PDI for 1960. The exceptions are Venezuela (see the explanation in note 10) and Turkey, whose 1960 PDI score of 59 reflects the fact that the military had taken control of the government during a crisis in 1960 and had temporarily suspended democratic liberties. Democracy was restored in 1961, however, as reflected in Turkey's 1965 PDI score of 76.

Among the set of countries classified as unstable democracies, all income-inequality measurements are either prior to or almost exactly contemporaneous with the breakdown of democracy. Otherwise, the inequality measurements are as close to the midpoint of the interval (1970) as possible.

Democracy was replaced internally by authoritarian rule during 1961–80 in 9 of the 33 countries with income inequality measurements for 1960–80. The typical method of replacing a democratic regime by an authoritarian regime is the military coup d'etat. Military-dominated authoritarian regimes replaced democracy in Brazil in 1964, in Panama and Peru in 1968, in Chile and Uruguay in 1973, and in Turkey in 1980. The executive coup, where the chief executive with the support of the military decides to retain power indefinitely, is a method of unconstitutional regime change similar to the military coup. An executive coup was carried out by President Ferdinand Marcos in the Philippines in 1972. Less dramatic but nevertheless significant regime change entailing a transformation from democracy to authoritarian rule also can occur through the imposition of long-term restrictions on political liberty. This kind of breakdown of democracy occurred in Malaysia in 1969, when parliament was suspended. It was restored in 1971, but with constitutional provisions for political and economic discrimination in favor of Malays and a ban on public and parliamentary discussion of these privileges and the concomitant restrictions on the rights of non-Malays. Finally, an exceedingly violent civil war between Christians and Muslims erupted in Lebanon in 1975, resulting in the demise of democracy and its replacement by a tenuous authoritarian rule bordering on anarchy.

The relationship between the upper-quintile income share[21] and democratic stability is shown in Table 6. The point-biserial correlation is significant at the .05 level and has a quite high magnitude ($-.80$). All democracies with high income inequality (an upper-quintile share of more than 55 percent) were unstable. These very inegalitarian democracies were highly susceptible to military coups, which were responsible for four of the six instances of instability. By contrast, slightly more than two-thirds (70 percent) of the democracies with an intermediate level of inequality (an upper-quintile share between 45 percent and 55 percent) maintained stability, and all of those with relatively low inequality (an upper-quintile share of less than 45 percent) were stable.

[21] The data are from Table 1 (T1), Table 5 (T5), and the World Bank (W).

Table 6. Stability of Democracy by Income Inequality

Upper 20% Income Share, 1960–80		Stability of Democracy, 1961–80	
(Source:Year)	%	Unstable	Stable
(T5)	69.2	Peru	
(T5)	62.1	Brazil	
(T1)	61.8	Panama	
(T5)	61.0	Lebanon	
(T1)	58.6	Turkey	
(W:70)	56.6	Malaysia	
(T1)	54.0	Philippines	
(T1)	54.0		Venezuela
(T1)	51.4	Chile	
(T1)	50.6		Costa Rica
(T1)	50.0		Trinidad
(T1)	49.2		India
(T1)	47.4	Uruguay	
(T1)	46.5		Italy
(T1)	46.4		France
(T1)	45.2		West Germany
(T1)	44.5		Sri Lanka
(T1)	44.0		Barbados
(T1)	43.0		Australia
(T1)	42.8		United States
(T1)	41.5		New Zealand
(T1)	41.0		Canada
(T1)	41.0		Japan
(T1)	40.0		Netherlands
(W:79–80)	39.9		Israel
(T1)	39.8		Belgium
(T1)	39.4		Ireland
(T1)	39.2		United Kingdom
(W:78)	38.0		Switzerland
(W:76)	37.5		Denmark
(T1)	37.3		Norway
(T1)	37.0		Sweden
(W:77)	36.8		Finland
	[$r = -.80$]		

Of course, the strong negative association between income inequality and democratic stability could simply be a reflection of a country's level of economic development. The argument that a high level of economic development is the principal cause of democratic stability has been the conventional wisdom in social science since Lipset (1959) first reported a positive association between stable democracy and a set of interrelated indicators of national wealth, industrialization, urbanization, education, and communications development.[22] Thus, if economic development

[22] According to Lipset (1981, pp. 469–76), this association has been repeatedly confirmed by subsequent research (e.g., Cutright 1963; Dahl 1971; McCrone and Cnudde 1967; Olsen 1968; Winham 1970). But Lipset ignores the fact that subsequent work has used either Cutright's (1963) measure of democracy, which confounds level and stability, or else measures of level of democracy only. Such analyses do not provide evidence relevant to the stability question.

does cause both income inequality and stable democracy, then the relationship between inequality and democratic stability might be spurious due to economic development.

Indeed, a strong inverse association exists for these democracies between level of economic development and the extent of income inequality. The correlation between ln Energy Consumption p.c., 1970 and the Upper 20 percent Income Share, 1960–80 is −.66. The point-biserial correlation between ln Energy Consumption p.c., 1970 and Stability of Democracy, 1961–80 is only +.58, however, while that between the Upper 20 percent-Income Share and Stability of Democracy is −.80.

Regressions of democratic stability on the Gini coefficient and the income shares of each of the five quintiles, controlling for economic development, are shown in Table 7. The Gini coefficient and the upper quintile income share are estimated to have negative effects on democratic stability, while the income shares of the lower four quintiles are estimated to have positive effects. The ln Energy Consumption p.c. variable is not significant at even the .10 level in any of these equations except that for the bottom quintile (7.6). Thus, level of economic development appears to be the irrelevant variable.

Figure 1 shows the irrelevance of economic development. Stable and unstable democracies are plotted by their joint scores on the Upper 20 percent Income Share and ln Energy Consumption p.c. variables. Countries above the dashed
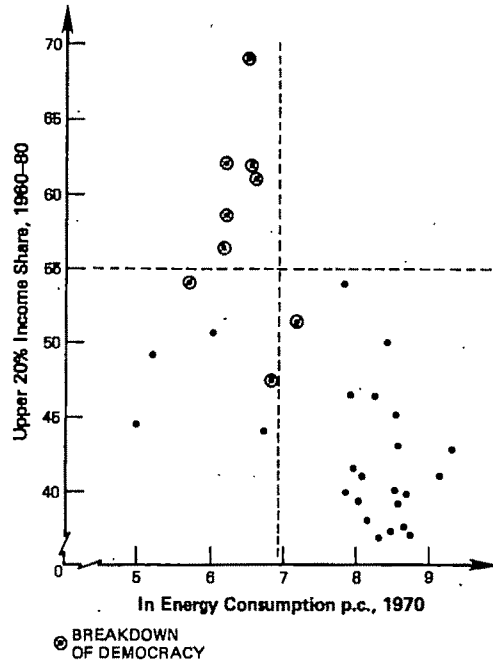
Fig. 1. Stability of Democracy by Income Inequality and Level of Economic Development.

horizontal line have extremely inegalitarian distributions of income; countries to the right of the dashed vertical line have a relatively high level of economic development (per capita energy consumption of more than 1,000).

The critical data points are contained in the

Table 7. Regressions[a] of Stability of Democracy on Gini Coefficient of Income Inequality and Quintile Income Shares, Controlling for Economic Development

| | Stability of Democracy, 1961–80 | | | | | |
	(7.1)	(7.2)	(7.3)	(7.4)	(7.5)	(7.6)
Intercept	1.71	2.24	−2.09	−1.26	−1.037	−1.24
ln energy consumption p.c., 1970	+0.65 (1.27)	+0.038 (0.70)	+0.064 (0.99)	+0.028 (0.47)	+0.066 (1.18)	+0.14** (3.04)
Gini income inequality, 1960–80	−3.90** (−5.12)					
Quintile 5 (81%–100%), 1960–80		−0.038** (−5.08)				
Quintile 4 (61%–80%), 1960–80			+0.11** (3.51)			
Quintile 3 (41%–60%), 1960–80				+0.12** (4.48)		
Quintile 2 (21%–40%), 1960–80					+0.12** (4.35)	
Quintile 1 (1%–20%), 1960–80						+0.16** (4.64)
Adj. R^2 =	.62	.62	.50	.58	.57	.59
(N =)	(33)	(33)	(33)	(33)	(33)	(33)

[a] t ratio in parentheses.
* $p \leq .10$, one-tailed.
** $p \leq .05$, one-tailed.

lower left-hand cell of the figure: democracies characterized both by a relatively low level of economic development and by income inequality scores that fall in the range of intermediate to relatively low. According to the economic development hypothesis, these democracies should have been unstable because of their relatively low level of development. By contrast, the income inequality hypothesis predicts that these democracies should have been stable because of their lack of extreme inequality. Since regime stability was maintained during 1960–80 by most less-developed democracies with intermediate or relatively low income inequality, the data indicate that income inequality is the predominant causal variable.

SUMMARY AND CONCLUSION

In testing the hypothesis that democracy reduces income inequality, the democracy concept can be defined operationally either as the level of democracy in a country at a single point in time or as a country's years of democratic experience. According to the level-of-democracy version of the hypothesis, a high level of democracy at any time is expected to produce a more egalitarian distribution of income. Thus, new democracies with few years of democratic experience are expected to be just as egalitarian as old democracies with many years of democratic experience. This is an implausible assumption. Since the causal influence of democratic institutions must operate through intervening mechanisms such as interest groups and political parties, it is more plausible to stipulate that the influence of democracy on income inequality operates gradually. The assumption of gradual causal influence requires a measure of years of democratic experience. According to the experience-of-democracy version of the hypothesis, a high level of democracy is expected to reduce inequality only if it is maintained for a relatively long time.

The evidence reported here, which is based on a larger sample of countries than in most prior studies, shows that the operational definition of democracy indeed makes a great deal of difference. The bivariate correlation between level of democracy and income inequality is much weaker than that between years of democracy and income inequality. Moreover, no significant inverse association obtains between level of democracy and income inequality, independent of a country's level of economic development. But when the level of democracy measure is replaced by continuous and qualitative measures of a country's experience of democracy, a significant negative effect of years of democracy on income inequality is observed, controlling for level of economic development and other potentially relevant determinants of inequality. These results support the argument that democracy has a negative influence on income inequality, one that operates gradually over time.

Of course, an effect of democratic institutions on the distribution of income must be transmitted through intervening causal mechanisms. The egalitarian political institutions of modern democracy provide all citizens with both the *opportunity* to participate in the governing process, as manifested by universal adult suffrage and free and fair elections, and the *opportunity* to contest governmental decisions, as manifested by rights of freedom of expression and association. Given the existence of an egalitarian political structure, it is plausible to expect that, over time, as the more numerous poorer members of the population organize into unions and other interest groups, and as parties of the social democratic left develop a solid electoral base, win seats in legislatures, and participate in or control the machinery of government, democracy becomes associated, in a *facilitative* sense, with a gradual reduction of economic inequality (see Lenski 1966, pp. 318–25). Thus, the influence of democratic political structure on income inequality should take the form of an indirect negative effect, operating through intervening processes such as the development of strong trade unions and socialist parties (see Stephens 1979). These transmission mechanisms that intervene between years of democracy and a reduction of income inequality are complex and await further research.

In regard to the hypothesis of a negative causal effect of income inequality on democracy, a measure of level of democracy is inappropriate because it is not sensitive to temporal variation in the inauguration and stability of democracy. If income inequality has a negative effect on democracy, it could operate either (1) by reducing the likelihood of the inauguration of democracy in countries under authoritarian rule; or (2) by causing a reduction of democracy in a country that was initially highly democratic, i.e., by causing a change from democracy to authoritarian rule. The genesis of democracy and the stability of democracy are separate questions requiring separate measures, and the causes of genesis may be quite different from the causes of stability (see Rustow 1970).

No support is found for the genesis hypothesis across a small sample of 23 countries (which, however, includes a majority of all countries that inaugurated democracy during the 1945–61 period). A more compelling theoretical argument can be made for the hypothesis that income inequality is incompatible with the stability of democracy than for the hypothesis that it is incompatible with the inauguration of democracy. When the stability question is

addressed on the basis of a sample that includes more than 80 percent of the relevant cases, a very strong inverse association is observed between income inequality and the likelihood of stability versus breakdown of democracy. Democracies as of 1961 with extremely inegalitarian distributions of income during 1960–80 all experienced a breakdown of democracy (typically due to a military coup d'etat), while a breakdown of democracy occurred in only 30 percent of those with intermediate income inequality. It did not occur at all among democracies with relatively egalitarian distributions of income. This negative effect of income inequality on democratic stability is independent of a country's level of economic development because, among democracies at relatively low levels of development, where one would expect instability if a positive relationship between economic development and democratic stability were the only causal force, regime stability was maintained by most democracies that did not have extremely inegalitarian distributions of income. Indeed, level of economic development, considered by many scholars to be the predominant cause of variation in the stability of democratic regimes, is found to be an irrelevant variable once income inequality is taken into account.

In sum, the evidence from a longitudinal perspective indicates that inverse causal relationships between political democracy and income inequality obtain in both directions. Democratic institutions, if maintained for a relatively long time, cause some gradual reduction of income inequality, independent of level of economic development. But if a democratic regime is inaugurated in a country with an extremely inegalitarian distribution of income, high inequality is likely to undermine the legitimacy of the regime and cause democratic institutions to be replaced by authoritarian rule. Thus, a high level of inequality will reduce a country's years of democratic experience. This vicious circle can be avoided if democracy is inaugurated (or restored) in a country with only intermediate or relatively low inequality. Otherwise, it seems that the only way to break out of this vicious circle is for a strong political party with redistributive goals to develop during the early years of democracy and then hold office for a sufficiently long period of time to implement policies that significantly reduce income inequality.

REFERENCES •

Adelman, Irma and Cynthia Taft Morris. 1967. *Society, Politics, and Economic Development*. Baltimore: Johns Hopkins University Press.

Ahluwalia, Montek S. 1974. "Income Inequality: Some Dimensions of the Problem." Pp. 3–37 in *Redistribution with Growth*, edited by Hollis Chenery, Montek S. Ahluwalia, C.L.G. Bell, John H. Duloy, and Richard Jolly. New York: Oxford University Press.

———. 1976. "Inequality, Poverty, and Development." *Journal of Development Economics* 3:307–42.

Bilson, John F.O. 1982. "Civil Liberty—An Econometric Investigation." *Kyklos* 35:94–114.

———. 1979. "Political Democracy and the Timing of Development." *American Sociological Review* 44:572–87.

———. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45:370–90.

Bollen, Kenneth A. 1983. "World System Position, Dependency, and Democracy." *American Sociological Review* 48:468–79.

Bollen, Kenneth A. and Burke D. Grandjean. 1981. "The Dimension(s) of Democracy: Further Issues in the Measurement and Effects of Political Democracy." *American Sociological Review* 46:651–59.

Bollen, Kenneth A. and Robert W. Jackman. 1985. "Political Democracy and the Size Distribution of Income." *American Sociological Review* 50:438–57.

Ballmer-Cao, Thanh-Huyen and Jürg Scheidegger eds. 1979. *Compendium of Data for World System Analysis*. Zürich: Soziologisches Institut der Universität.

Bornschier, Volker and Christopher Chase-Dunn. 1985. *Transnational Corporations and Underdevelopment*. New York: Praeger.

Coleman, James S. 1960. "Conclusion: The Political Systems of the Developing Areas." Pp. 532–76 in *The Politics of the Developing Areas*, edited by Gabriel A. Almond and James S. Coleman. Princeton: Princeton University Press.

Cotler, Julio. 1978. "A Structural-Historical Approach to the Breakdown of Democratic Institutions: Peru." Pp. 178–205 in *The Breakdown of Democratic Regimes: Latin America*, edited by Juan J. Linz and Alfred Stepan. Baltimore: Johns Hopkins University Press.

Cutright, Phillips. 1963. "National Political Development: Measurement and Analysis." *American Sociological Review* 28:253–64.

Comisión Económica para América Latina. 1983. "Notas sobre la Evolucion del Desarrollo Social del Istmo Centroamericano hasta 1980." (Notes on the Evolution of Social Development in the Isthmus of Central America Since 1980.) Pp. 93–126 in *Centroamerica: Indicadores socioeconómicos para el Desarollo*, by R. Ramalinga Iyer, Gonzalo Ramirez, Carlos Raabe, Guillermo Molina Chocano, Sergio Reuben, et al. San José, Costa Rica: Ediciones Flacso.

Dahl, Robert A. 1971. *Polyarchy*. New Haven: Yale University Press.

Encyclopedia Britannica. 1966. vol. 22. Chicago: William Benton.

Finer, S.E. 1962. *The Man on Horseback: The Role of the Military in Politics*. New York: Praeger.

Gastil, Raymond D. 1986. *Freedom in the World, 1986*. Westport, CT: Greenwood Press.

van Ginnekan, Wouter. 1976. *Rural and Urban Income Inequalities*. Geneva: International Labour Office.

Hewitt, Christopher. 1977. "The Effect of Political Democracy and Social Democracy on Equality in Industrial Societies: A Cross-National Comparison." *American Sociological Review* 42:450–64.

Jackman, Robert W. 1973. "On the Relation of Economic Development to Democratic Performance." *American Journal of Political Science* 17:611–21.

———. 1974. "Political Democracy and Social Equality: A Comparative Analysis." *American Sociological Review* 39:29–45.

Jain, Shail. 1975. *Size Distribution of Income*. Washington, DC: World Bank.

Kline, Harvey F. 1985. "Colombia: Modified Two-Party and Elitist Politics." Pp. 249–70 in *Latin American Politics and Development*, 2nd ed., edited by Howard J. Wiarda and Harvey F. Kline. Boulder, CO: Westview.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45:1–28.

———. 1963. "Quantitative Aspects of the Economic Growth of Nations, VIII: The Distribution of Income by Size." *Economic Development and Cultural Change* 11:1–80.

LaFeber, Walter. 1979. *The Panama Canal,* expanded ed. New York: Oxford University Press.

Lecaillon, Jacques, Felix Paukert, Christian Morrisson, and Dimitri Germidis. 1984. *Income Distribution and Economic Development*. Geneva: International Labour Office.

Lenski, Gerhard. 1966. *Power and Privilege*. New York: McGraw-Hill.

Levine, Daniel H. 1978. "Venezuela Since 1958: The Consolidation of Democratic Politics." Pp. 82–109 in *The Breakdown of Democratic Regimes: Latin America,* edited by Juan J. Linz and Alfred Stepan. Baltimore: Johns Hopkins University Press.

Lijphart, Arend. 1977. *Democracy in Plural Societies*. New Haven: Yale University Press.

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53:69–105.

———. 1981. *Political Man,* expanded ed. Baltimore: Johns Hopkins University Press.

Luzuriaga C., Carlos and Clarence Zuvekas, Jr. 1983. *Income Distribution and Poverty in Rural Ecuador, 1950–1979*. Tempe, AZ: Center for Latin American Studies, Arizona State University.

Lydall, Harold. 1979. *A Theory of Income Distribution*. New York: Oxford University Press.

McCrone, Donald J. and Charles F. Cnudde. 1967. "Toward a Communications Theory of Democratic Political Development." *American Political Science Review* 61:72–79.

McDonald, Ronald H. 1985. "El Salvador: The Politics of Revolution." Pp. 528–44 in *Latin American Politics and Development*, 2nd ed., edited by Howard J. Wiarda and Harvey F. Kline. Boulder, CO: Westview.

Martz, John D. 1968. "Venezuela." Pp. 199–230 in *Political Forces in Latin America: Dimensions of the Quest for Stability,* edited by Ben G. Burnett and Kenneth F. Johnson. Belmont, CA: Wadsworth.

Moore, Barrington. 1966. *The Social Origins of Dictatorship and Democracy*. Boston: Beacon.

Nordlinger, Eric A. 1977. *Soldiers in Politics: Military Coups and Governments*. Englewood Cliffs, NJ: Prentice-Hall.

Olsen, Marvin E. 1968. "Multivariate Analysis of National Political Development." *American Sociological Review* 33:699–712.

Paukert, Felix. 1973. "Income Distribution at Different Levels of Development: A Survey of Evidence." *International Labour Review* 108:97–125.

Peterson, Robert L. 1968. "Guatemala, Honduras, El Salvador, and Nicaragua." Pp. 57–90 in *Political Forces in Latin America: Dimensions of the Quest for Stability,* edited by Ben G. Burnett and Kenneth F. Johnson. Belmont, CA: Wadsworth.

Powell, G. Bingham, Jr. 1982. *Contemporary Democracies*. Cambridge: Harvard University Press.

Rubinson, Richard and Dan Quinlan. 1977. "Democracy and Social Inequality: A Reanalysis." *American Sociological Review* 42:611–23.

Russett, Bruce M. 1965. *Trends in World Politics*. New York: Macmillan.

Rustow, Dankwart A. 1967. *A World of Nations*. Washington, DC: Brookings.

———. 1970. "Transitions to Democracy." *Comparative Politics* 2:337–63.

Sartori, Giovanni. 1976. *Parties and Party Systems*. New York: Cambridge University Press.

Sawyer, Malcolm. 1976. "Income Distribution in OECD Countries." *OECD Economic Outlook: Occasional Studies,* 3–36.

Schlesinger, Stephen and Stephen Kinzer. 1983. *Bitter Fruit*. Garden City, NY: Anchor Press/Doubleday.

Snyder, David and Edward Kick. 1979. "Structural Position in the World-System and Economic Growth, 1955–70: A Multiple Network Analysis of Transnational Interactions." *American Journal of Sociology* 84:1096–1126.

Stack, Steven. 1978. "Internal Political Organization and the World Economy of Income Inequality." *American Sociological Review* 43:271–72.

Stephens, John D. 1979. *The Transition from Capitalism to Socialism*. London: Macmillan.

Stevens, Evelyn P. 1985. "Mexico in the 1980s: From Authoritarianism to Power Sharing?" Pp. 403–45 in *Latin American Politics and Development,* 2nd ed., edited by Howard J. Wiarda and Harvey F. Kline. Boulder, CO: Westview.

Taylor, Charles L. and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators,* 2nd ed. New Haven: Yale University Press.

Weede, Erich. 1980. "Beyond Misspecification in Sociological Analysis of Income Inequality." *American Sociological Review* 45:497–501.

———. 1982. "The Effects of Democracy and Socialist Strength on the Size Distribution of Income." *International Journal of Comparative Sociology* 23:151–65.

Weede, Erich and Horst Tiefenbach. 1981a. "Some Recent Explanations of Income Inequality: An Evaluation and Critique." *International Studies Quarterly* 25:255–82 and 289–93.

———. 1981b. "Correlates of the Size Distribution of Income in Cross-National Analyses." *Journal of Politics* 43:1029–41.

Weinstein, Martin. 1975. *Uruguay*. Westport, CT: Greenwood Press.

Welch, Claude E. and Arthur K. Smith. 1974. *Military Role and Rule: Perspectives on Civil-Military Relations*. North Scituate, MA: Duxbury Press.

Wesson, Robert and David V. Fleischer. 1983. *Brazil in Transition*. New York: Praeger.

Wiarda, Iêda Siqueira. 1985. "Venezuela: the Politics of Democratic Developmentalism." Pp. 293–316 in *Latin American Politics and Development*, 2nd ed., edited by Howard J. Wiarda and Harvey F. Kline. Boulder, CO: Westview.

Winham, Gilbert R. 1970. "Political Development: A Causal Model." *American Political Science Review* 64:810–18.

World Bank. 1980, 1981, 1982, 1983, 1984, 1985. *World Development Report*. New York: Oxford University Press.

THE POLITICAL GEOGRAPHY OF BELGIAN FASCISM:
THE CASE OF REXISM*

WILLIAM BRUSTEIN

University of Utah

Though Italy and Germany furnish the best known cases of fascism's political geography between the two World Wars, one of the more·intriguing and less known examples of regional variation in fascist strength occurred in Belgium. In the 1936 Belgian legislative elections, the fascist Rexist movement captured over 30 percent of the popular vote in the province of Luxembourg, compared to only 9·percent in Hainaut. This study applies a general theory of regionalism to the case of Belgian fascism. The data support the paper's hypothesis that divergent regional patterns of Belgian voting in 1936 can be largely ascribed to the existence of particular modes of production that were responsible for distinctive regional constellations of interests.

The rise of European fascism during the interwar period has elicited a voluminous literature. Though quite helpful at explaining the general causes of fascist successes in Europe, the best-known explanations of fascism fail to explain locational variation in support for fascism:[1] Lipset's (1981) and de Felice's (1966) lower-middle-class theory, Kornhauser's (1959) and Parsons' (1954) mass-society theory, Nolte's (1966) and Moore's (1966) resistance to modernization theory, Fromm's (1966) and Reich's (1972) pathological psychological theory, and Palme Dutt's (1974) and Poulantzas' (1974) last stage of capitalism theory. None of these theories adequately explains why, in the Reichstag elections between 1930 and 1933, the average Nazi vote was 40 percent in Schleswig-Holstein and 22.5 percent in Lower Bavaria (Pollack 1944, p. 90) or, why, in the May 1921 Italian Parliamentary elections, 4 of the 10 deputies elected in Venezia Giulia were fascists, while none of the 52 deputies in Sicily was fascist (Petersen 1975, p. 643).

In contrast to the highly general treatments of fascism, several case studies of the rise of European fascist movements appear to hold promise as explanations of fascism's political geography. Much of the literature on German fascism calls attention to the differences in confessional faiths in Germany during the 1920s and 1930s. They argue that Protestant regions voted heavily for the Nazi party while Catholic regions voted heavily for the Catholic Center Party or the Bavarian People's Party.[2] The results of the 1930 and 1932 Reichstag elections confirm that the heavily Protestant North and East voted substantially for the Nazis while the predominantly Catholic South and West voted equally heavily for the Catholic parties. However, the confessional faith explanation cannot account for variation in fascist support in countries where religious differences were absent. Thus, although in both Italy and Belgium the Catholic church maintained a religious hegemony, significant regional variation occurred in support of Mussolini's fascist party and Degrelle's Rexist party. Even in Germany, numerous exceptions appeared, where heavily Protestant areas diverged dramatically in their support of the Nazi party. For example, in the Baden Landtag election of 1929, Hitler obtained 10 percent of the vote in the solidly Protestant village of Kurnbach and 64 percent of the vote in the equally Protestant village of Zausenhausen just four kilometers away (Faris 1975, p. 167).

The literature on Italian fascism generally attributes fascist popularity to a post-World·War I reaction to socialism. Thus, Snowden (1972), Corner (1975), and Cardoza (1982) argue that between 1918 and 1922 fascism's greatest success occurred in areas that voted overwhelmingly for the Italian Socialist party. Fearing an· eventual socialist takeover, much of the Italian urban and rural population turned to Mussolini's fascists as a last defense. And, indeed, the evidence demonstrates that in the socialist strongholds of Lombardy, Emilia, Tuscany, and Apulia, fascism received substantial support,

---

[1] For a very helpful summary of the strengths and weaknesses of these general explanations of the rise of ·fascism, see Eley (1983).

•

[2] The use of divergent confessional faiths to explain the political geography of German Nazism can be found in such works as Pridham (1973), Passchier (1980), and Hamilton (1982).

while it made few inroads in such regions as Abruzzi, Molise, Sicily, Basilicatà, and Calabria, where socialism failed to establish a foothold. Though fascist popularity as a reaction to socialism may be applicable to Italy, it fails to account for the political geography of fascism elsewhere. For example, it cannot explain why the Nazi party did so well in Schleswig-Holstein, Saxony, and Franconia, areas where leftist parties had relatively minor support, while the Nazi party did less well in the Ruhr and Rhine, where the German left had built a sizable following.

In brief, explanations of the rise of fascism drawn from case studies of the German or Italian experience offer considerable promise in accounting for the political geography of fascism within both countries. However, these explanations are not generalizable beyond their specific case. This paper applies a general theory of mode of production to explain the political geography of fascism in Belgium.

In the 1936 legislative elections, the fascist Rexist movement garnered 31.1 percent of the popular vote in the French-speaking Luxembourg province, compared to 9.4 percent in the French-speaking Hainaut province. A history of Belgian fascism should kindle our interest for two reasons.[3] First, in the 1936 election fascists (Rexists and Flemish nationalists) stunned the Belgian electorate by winning 37 of 202 parliamentary seats. This feat was remarkable since the newly founded Rexist movement (which won 31 parliamentary seats) had no party organization or prior legislative experience (Hojer 1946, p. 248). Second, Belgium seemed so unsuitable for fascism. Juan Linz (1976, p. 7) has noted that fascism is usually seen as a novel response to crises brought on by such post-war dislocations as defeat, ambivalence about a nation's entry into the war, disappointment with the peace terms, or unsuccessful revolutionary attempts. But Belgium emerged as one of the victors in World War I and escaped the deprivation and humiliation associated with the post-war period. Moreover, Belgium had a stable and well-established parliamentary system, no tradition of indigenous right-wing groups, and minerally-rich central African colonies (Chertok 1975, p. 1).[4] Eugene Weber

(1964) has described this apparent mismatch: "(Belgium) a country whose problems were in no way dramatic and whose people, solid and often stolid, inclined neither to excesses nor to histrionics" (p. 122).

How are we to explain the political geography of Belgian fascism? In earlier studies (Brustein 1981, 1988), I presented a general theory of regionalism and applied it to France, arguing that the persistence of political regionalism is directly attributable to the existence of distinctive modes of production in the Western and Mediterranean regions. These regional modes of production have produced different constellations of interests which, in turn, have led to right-wing voting in Western France and left-wing voting in Mediterranean France. To increase the validity of my theory, I hope to demonstrate that the political geography of Belgian fascism can equally be ascribed to the presence of different modes of production.

THE "MODE-OF-PRODUCTION" MODEL OF VOTING

Marx ([1859] 1959, p. 43) used the term *mode of production* to describe a combination of productive forces (social and technical means by which production is organized and carried out) and social relations (the institutions and practices associated with the way goods are produced, exchanged, and distributed). To employ the mode-of-production concept in analysis, suitable empirical indicators of productive forces and social relations are needed. Economic activity (forms of individual economic production) and property rights (the conditions or manner of holding property) are measurable proxies of these two key dimensions of mode of production.

Following Marx, the combination of economic activity and property rights forms the economic base of society and shapes its social characteristics. Such factors as patterns of settlement (population concentration and town-

---

[3] Scholars debate whether Rexism is really fascism. For Étienne, Rexism is not fascism but rather a Catholic, nationalistic, authoritarian conservative movement that only developed into a true fascist movement after 1937. By contrast, Stengers (1965) and Carpinelli (1973) see Rexism as fascism in both form and content.

[4] The 1936 Belgian election occurred at a time of considerable turmoil. The Belgian economy was in the midst of a depression. Between 1930 and 1935, unemployment jumped from 42,000 to 355,000 (Goris

1946, pp. 66–81). In 1935, the Belgian government was forced to devalue the franc to 1/10 of its pre-war gold value (Fitzmaurice 1983, pp. 40–41). Also, several major financial institutions including the Catholic's Farmer Corporation (Boerenbond) and the Banque du Travail went bankrupt (Chertok 1975, p. 85; Goris 1946, pp. 73–76). The political situation was no better. Between 1919 and 1936, one coalition government after another resigned. Political legitimacy suffered from numerous public disclosures of politico-financial scandals involving top government and banking officials. Persisting social problems continued to cast long shadows over the Belgian landscape as well. Two of the more significant controversies surrounded the defense of the Free Catholic schools and minority status of the Flemish language (Hojer 1946, pp. 54–59).

countryside association) and class composition (relative presence or absence of various classes) are affected by the mode of production, suggesting that it can be used to explain the various forms of patterns of settlement and class composition in different societies.[5] For example, in many parts of the world, a commercialized agriculture accompanies short-term cash tenancy or agricultural wage labor, nucleated villages, intense town-countryside association, and a heterogeneous population. Alternatively, a subsistence agriculture accompanies hereditary tenures or agricultural servant labor, dispersed homesteads, weak town-countryside association, and a homogeneous population. Why, for example, should agricultural servant labor rather than wage labor be associated with a subsistence cultivation? According to the theory, this association follows because each component of the mode of production influences other social components in the social unit. In the example above, farm servant labor is more economical in a subsistence economy because it solves the problem of a farm's cash flow by delaying money payments for work done until after the harvest and by paying much of the real wage in food, drink, and lodging.

The mode of production affects voting both directly and indirectly. People's perceptions of their interests are derived directly from their economic activity and property rights, and determined indirectly by the dominant structure of their economic activity and property rights. Workers' economic activity and property relations affect the stance they take on major political issues. For example, market-oriented producers anticipate different personal economic consequences of protective tariffs than do subsistence producers, and property owners see a different set of benefits and damages resulting from rent control than do tenants. It follows that market-oriented producers, subsistence producers, property owners, and tenants should support the political programs that best represent their interests.

The dominant structure of economic activity and property rights also conditions the particular patterns of settlement. More intense town-countryside association is expected in regions with market-oriented economic activity than in those with subsistence economic activity. Patterns of settlement have an independent effect on workers' perceptions of their interests. First and most importantly, patterns of settlement

influence which individuals and groups in a community control the allocation of resources. To the extent that people need these resources and wish to maintain uninterrupted access to them, they will depend on their suppliers and have a stake in their welfare. Consequently, they will view political programs that threaten the welfare of their supplier as threats to their own livelihood and will be unlikely to oppose the programs their suppliers support.

Second, patterns of settlement influence the availability of information and play an important role in groups' awareness of their interests. Isolated communities should be expected to have fewer sources of information than nonisolated communities. If people depend on a particular source of information, be it newspapers, books, or associates, their interests and their voting are likely to reflect the source's formulation of their interests.

In short, the mode of production affects political preferences by shaping specific interests on which people base their action. In the case of Belgium, my theory hypothesizes that, where the Rexist party program best reflected local interests, it should have received strong popular support; where other party programs better reflected cultivators' interests, Rexism should have received weak popular support. My study focuses on Wallonia, the French-speaking half of Belgium, since Rexism was principally confined to this half of Belgium. Flanders, the Flemish half, had its own fascist movement, the Vlaams National Verbond (VNV).

I first examine the various agricultural modes of production of Wallonia. Because the issue of regional variation in fascist popular support is most dramatically raised by two of Wallonia's principal agricultural provinces, Hainaut and Luxembourg, I concentrate on agricultural modes of production. I will then discuss the political programs of Belgium's principal parties during the interwar period and test the mode of production model for the 1936 Belgian legislative election.

REGIONAL MODES OF
PRODUCTION IN WALLONIA

Though Belgium was one of Europe's most industrialized nations during the 1930s, agriculture employed a sizable proportion of its population. Nearly 60 percent of Belgium's total land surface was devoted to agriculture. Farms were generally small, averaging between one and one-half to two hectares. Owner-operated farms made up 48 percent of the total number, while tenant farms comprised 52 percent (Misner 1946, p. 148). Climate and soil type largely determined regional agricultural specialization. Wallonia encompasses four major

---

[5] My use of mode of production differs from that of Marx ([1859] 1959). In a strict Marxian sense, I am simply specifying variants of the capitalist mode of production. I employ this strategy because I find Marx's use of the concept ambiguous and insensitive to the enormous variation within modes of production.

72   AMERICAN SOCIOLOGICAL REVIEW

agricultural regions (Misner 1946; Henry 1937; Quicke 1966): the sandy loam, loam, limestone, and Ardennes regions.

The sandy loam region includes the southeastern half of Brabant province and the northern edge of Hainaut and Namur provinces. Such cities as Brussels, Louvain, and Tirlement belong to this region. The sandy loam region is one of the richest agricultural regions in Belgium and, while polyculture is widespread, market-gardening is the principal agricultural activity (Henry 1937, p. 9). Superior yields and proximity to urban markets make this region's agriculture highly market-oriented (Misner 1946, p. 159; Quicke 1966, pp. 79–84). Approximately two-thirds of the agricultural surface is under tenancy and one-third is under owner-cultivation. Nearly three-quarters of all farms are less than one hectare (Henry 1937, pp. 12, 16). Population agglomeration in this region averages between 490 and 1,700 inhabitants per km², making it among the most densely populated regions in Belgium (*Annexe Statistique* 1951).

The loam region of central Belgium covers Hainaut province and small sections of Namur and Liége province. The soil is deep and contains a productive mixture of clay, sand, and calcareous material creating Belgium's most productive agriculture (Misner 1946, pp. 60–61). The region specializes in the cultivation of numerous grains, market-gardening and forage crops, and industrial plants, especially sugarbeets (Henry 1937, p. 28; Quicke 1966, pp. 63–69; Misner 1946, p. 162). This region possesses a highly market-oriented agriculture because of its superior yields and its proximity to urban markets (Misner 1946, p. 161). Three-fourths of the agricultural surface is under tenancy with the remaining one-fourth under owner-cultivation (Henry 1937, pp. 12, 16). Like the sandy loam region, most farms here are less than one hectare. Tenants work most of these small farms, while industrial workers own and operate

Fig. 1. Belgian Provinces

others (Misner 1946, p. 160). But large farms also exist in the loam region. Misner (1946, p. 161) notes that most of Belgium's large farms are located here. These farms frequently employ agricultural wage labor. Population agglomeration varies greatly in this region, with the southeastern zone averaging between 45 and 119 inhabitants per km² and the central zone averaging between 370 and 899 inhabitants per km² (*Annexe Statistique* 1951).

The limestone region is situated principally in the provinces of Namur and Liége, but includes a small area within the eastern part of Hainaut province. Its soil contains considerable limestone, but is nonetheless quite productive. The region specializes in cattle-raising, but it also produces some grains, alfalfa, red clover, and nursery products (Quicke 1966, pp. 45–47; Misner 1946, p. 162). This region's agriculture is also highly market-oriented for the same reasons found in the sandy-loam and loam regions, that is, high agricultural yields and proximity of urban markets (Misner 1946, p. 162). Again, like the sandy-loam and loam regions, small-tenant farming predominates in the limestone region (Henry 1937, pp. 16, 28; Misner 1946, p. 162). Population agglomeration averages from 70 to 179 inhabitants per km² (*Annexe Statistique* 1951).

The Ardennes region covers almost all of the province of Luxembourg and the southeastern third of Liége province. This agricultural region differs enormously from the other three. With an average elevation between 400 and 700 meters, this region is the highest in Belgium and the rainiest and coldest as well. The land is principally covered by forests. The soil is less fertile than elsewhere in Wallonia. The cultivation of crops is secondary; the principal economic activity is cattle-raising (Misner 1946, p. 163; Quicke 1966, pp. 35–45; Henry 1937, p. 28). More than in any of the other regions of Wallonia, the Ardennes' economic activity is geared toward subsistence, probably because of its low yields and distance from urban markets (Misner 1946, p. 163). While small farms and tenancy predominate elsewhere in Wallonia, medium farms and owner-cultivation prevail in the Ardennes (Henry 1937, pp. 12, 16, 27). Population is relatively dispersed, as measured by an average population agglomeration of 30 to 69 inhabitants per km² (*Annexe Statistique* 1951).
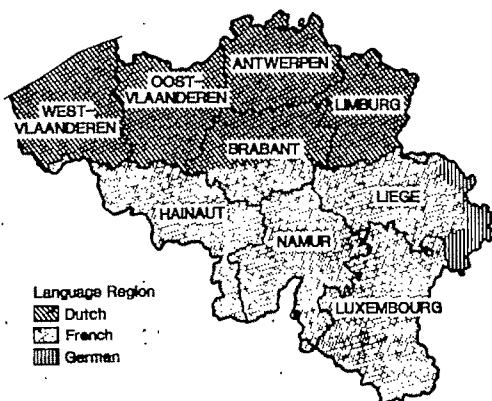
POLITICAL PARTIES AND ISSUES IN BELGIUM DURING THE INTERWAR PERIOD

During much of the interwar period, the Belgian political landscape was divided into right, center, and left. The Catholic Union stood on

the right, holding more than one-third of the Belgian electorate, the Liberal party held the center maintaining roughly one-sixth of the electorate, and the Socialists or Belgium Workers Party stood on the left with a little more than one-third of the electorate (Hojer 1946, p. 54).[6]

The Catholic Union represented the interests of the Church. It strongly advocated governmental support for Catholic schools and the right of the Church to run its schools (*écoles libres*). The party was pro-monarchy and viewed the well-run society as one that was hierarchically ordered and inspired by God. The Catholic Union also favored a strengthening of local and provincial privilege at the expense of the central government. The party also supported laws to protect private property and the family. However, the Union's liberal wing (Ligue Nationale des Travailleurs Chrétiens), though accepting the party's conservative positions on private property, religious issues, and local rights, supported the left's positions on many economic issues such as the eight-hour workday, minimum salary, affordable housing, social insurance, and a lowering of indirect taxes.

The Liberal Party's program was characterized by its long-standing policy of anticlericalism. This was most apparent in its unceasing drive to keep the Catholic church out of education. The Liberals saw themselves as proponents of toleration and enlightenment which, they argued, could only be achieved through an educational system free of all dogma. On economic issues the party defended the free-enterprise system. On the one hand, the Liberals supported the left's policies of an eight-hour workday, collective bargaining, social insurance, affordable housing, and a progressive taxation on income and inheritance. On the other hand, the Liberals opposed the left's plan to curb the free-enterprise system and the Communist Party's class struggle to reshape society. The Liberals also favored a strong military and opposed suffrage for women (Hojer 1946, pp. 46–48).

The Socialists advocated a reformist rather than revolutionary strategy to achieve workers' goals. According to party leaders, power could be attained through the electoral process. On religious issues, the Socialists, like the Liberals, advocated reducing the Catholic church's influence over educational matters. On economic matters, the Socialists were the principal advocates of lower indirect taxes, social insurance, an eight-hour workday, a minimum salary, affordable housing, and restrictions on the free-enterprise system. The Socialist Party

was the only major party to directly address the concerns of small tenant farmers, which is of particular importance to this study. The party called for legislation to provide greater security of tenure to these farmers. In contrast to the Liberals' pro-military stance, the Socialists proposed disarmament (until 1933) and a reduction of the military service requirement to six months (Hojer 1946, pp. 50–51).

The relative stability of Belgium's tripartite political landscape was severely shaken in 1936 with the unprecedented electoral success of the Rexist party. The Rexists emerged from the ranks of the Catholic Union. Accusing the Catholic Union of weakness and inactivity in the face of a corruption-ridden society, Rexist founder Léon Degrelle and his followers promised bold measures to restore order. At the heart of the Rexist program was a call for a corporate state modeled on fascist Italy (Chertok 1975, pp. 77–78). According to Léon Degrelle, corporatism was the best means to overcome the chaos of class struggle. The Rexist corporate state would be authoritarian and fully imbued with Christian values. As its first political act, the Rexist state would carry out a physical and moral reform of the Belgian nation (Denis 1936, p. 106; Scheppens 1980, p. 507; Étienne 1968, pp. 90–95).

The Rexists opposed the parliamentary system and called for the complete elimination of political parties. Rexists felt that political parties had sown national discord and their leaders were to blame for the numerous politico-financial scandals that had riddled the country during the 1930s (Chertok 1975, pp. 86–88; Étienne 1968, pp. 45–49).

On economic issues, the Rexists were decidedly anticommunist, seeing communism as the chief destructive force in the world (Weber 1964, p. 125; Schepens 1980, p. 507; Denis 1936, p. 106). They were equally opposed to big business; they blamed the major financial institutions for the worldwide economic depression and the impoverishment of small- and medium-scaled family-run businesses (Étienne 1968, pp. 90–95; Chertok 1975, pp. 86–88). But they were not opposed to private property or to capitalism, and they demanded that the state aid small- and medium-sized businesses and farms (Carpinelli 1973, p. 78; Étienne 1968, pp. 90–95; Wallef 1980, pp. 520–21). In particular, they called for more accessible agricultural credit and restrictions on large agro-businesses (Daye 1937; Degrelle 1938, pp. 107–9; Chertok 1975, p. 103).

One key Rexist concern was the plight of the family. For them, the family was the core of society, and they felt the state should do everything possible to protect, favor, and strengthen the family (Denis 1936, p. 106;

---

[6] The remaining 13 percent of the electorate was usually divided among extreme right-wing and left-wing parties.

Schepens 1980, p. 507; Étienne 1968, pp. 90–95). They believed the state should protect the family from such evils as pornography and prostitution, guarantee inheritance based on direct descendancy, and replace universal suffrage by the "integral family vote," which would have given a second vote to families with at least four children (Denis 1936, p. 106; Schepens 1980, p. 507; Étienne 1968, pp. 90–95).

WHO VOTED REX?

. Rexism should have received strong support in those regions where its program most closely reflected the interests of cultivators. Furthermore, the mode of production in these regions should most likely have been marked by subsistence economic activity, small- and medium-sized owner-occupied farms, and population dispersion. Rexism should have received weak support in those regions where the program of other political parties better represented the interests of cultivators. These regions should have possessed a mode of production marked by commercial economic activity, small tenant farms or large estates with agricultural wage laborers, and high population concentration.

Among Belgium's political parties, the Rexist party was the strongest proponent of large families. Because it called for the replacement of universal suffrage by the "integral family vote," voters in regions where large families were predominant should have been inclined to favor the Rexists. Further, these regions should most likely have had a subsistence economic activity and population dispersion.

Families generally tend to be larger in subsistence than in market-oriented communities because, in subsistence communities, economic activity is geared to family use, and large families represent an important economic asset in cheap labor (Lenski and Lenski 1982, p. 183). However, in market-oriented communities, production is geared to exchange, and large families are an economic liability. Belgian marital-fertility data for 1890 to 1910 support this argument. During this period, the average marital fertility in the subsistence-oriented Ardennes region was .530, while in the more market-oriented regions it was .343 for the sandy loam, .305 for the loam, and .372 for the limestone (Lesthaeghe 1977, p. 106).[7]

---

[7] This finding is further corroborated by the overall fertility by province for 1930. Overall fertility for the relatively subsistence-oriented province of Luxemburg was .203 compared to .145 for Brabant, .155 for Hainaut, .149 for Liége, and .178 for Namur (Lesthaeghe 1977, p. 130).

Moreover, we should expect to find larger families in communities of high population dispersion than in communities with high population concentration. In dispersed habitats, great distances make finding reliable farm labor difficult and expensive. The need for cheap farm labor stimulates having additional children.

Among the competing political parties, the Rexists most strongly favored the small- and medium-sized family-owned farm. Since the Rexists presented themselves as the saviors of the small- and medium-sized family-owned farm, we should expect to find considerable support for the Rexist party in regions with these farms. By contrast, the small tenant farmers and salaried agricultural laborers should have been more inclined to favor the Socialist Party because the Socialists advocated an improvement in tenant leases and a minimum wage for laborers.

Central to the Rexist political doctrine was the corporatist view of society. They believed society resembles a good patriarchal family; it should be hierarchical, with the elites in control. What kinds of people might have favored such a view of society? People who were totally dependent on big landowners for vital resources were likely to favor a society in which these landowners maintain special privileges. The recipients of a resource are interested in securing an uninterrupted flow of that resource. If these big landowners are the sole providers of vital resources, then it is in the material interests of the recipients to support the welfare of these landowners. People residing in regions marked by high population dispersion should be most apt to support the welfare of the big landowners. The great distances between cultivators makes social interaction and the emergence of peasant organizations more difficult. In these regions, resource allocation usually falls into the hands of big landowners. In contrast, a relatively concentrated population facilitates social interaction, lowering a group's costs of organization. Big landholders cannot monopolize resources; with alternative suppliers of comparable resources, groups are less likely to be dependent on any one source.

People residing in regions characterized by a subsistence economic activity should also be likely to support the corporatist view of society. Groups in a subsistence economy are less likely than those in a market-oriented economy to have surpluses to fall back on during crises. Consequently, they are likely to become dependent on the big landholders for supplies during crises.

Thus, Rexism should have received more votes in areas where the mode of production engendered a subsistence economic activity, small- and medium-sized farms, and population dispersion. In these regions, the constellation of

peoples' interests should have corresponded closely to the Rexist party program. By contrast, Rexism should have obtained fewer votes in areas where the mode of production engendered a market-oriented economic activity, small tenant farms or farms employing agricultural labor, and population concentration. In these regions, the constellation of peoples' interests should have coincided with the programs of other political parties.

While the theory being tested argues that people's perceptions of their interests and their resulting voting patterns are shaped by the mode of production, there are undoubtedly other factors that had an impact on the political geography of Rexism. One factor was party organizational capacity, which varied from region to region. Parties often maintained extensive networks and effective organizations in one region and not in others. For instance, because the Nazis encountered tremendous difficulty developing party networks outside of Bavaria during the early 1920s, their early electoral successes (albeit minor) were confined to that area (Pridham 1973). We can expect that voting should have reflected the extent to which a party could reach voters. However, in the case of Belgian Rexism in 1936, party organizational capacity does not appear to be a major issue since the party had no established networks in any agricultural region (Hojer 1946, p. 248). The Rexist movement relied principally on its numerous publications and public meetings to disseminate its program.

A second possible influence on Rexist strength may have been the traditional Catholic conservatism of particular Belgian regions. Rexism's greatest electoral victories occurred in provinces where Catholic conservatism was strong. It appealed to traditional Catholics because it advocated state subsidies to Catholic schools and to large families. Moreover, for many voters in 1936, the Rexist party, rather than the Catholic Union, was perceived to be a more forceful voice for traditional Christian values. I speculate, however, that the conservatism of these regions, whether it was expressed in terms of traditional Catholic conservatism or Rexism, was conditioned by the mode of production in these regions.

A third factor may have been the effects of border nationalism. Linz (1976) noted that the Italian fascist success in Venezia Giulia between 1919 and 1922 can be partially explained by strong border nationalism in the province because of its proximity to Austria and Yugoslavia. Similarly, Rexism won more votes in the border provinces of Luxembourg and Liége, where history and geographical location had nurtured strong anti-German sentiments (Linz 1976, p. 92; Stengers 1965, p. 145; Étienne

1968, pp. 40–42). Rexism appealed to people in these border regions because it called for a Greater Belgium and for vigilance toward Germany (Stengers 1965, p. 39; Étienne 1968, pp. 40–42).

A fourth factor may have been that Léon Degrelle, the founder of the Rexist movement, was born in the town of Bouillon in Luxembourg province. Some people may have voted for the Rexists because the founder was a native son. However, this factor cannot explain Rexism's strong electoral performance in the provinces of Namur and Liége.

A fifth influence may have been the effects of the depression. The 1936 election occurred at a time of considerable economic turmoil. In particular, the collapse of the Boerenbond and the Banque du Travail had brought considerable distress to the rural middle class, which was relatively large in the predominantly rural provinces of Luxembourg, Liége, and Namur (Goris 1946, pp. 73–76). Rexism's frequent criticism of the large banks and agro-businesses may have especially appealed to the distressed rural middle class in these provinces. However, the economic crisis of the mid-1930s cannot explain Rexism's relatively strong electoral performance in these same provinces in the 1939 election, when the economic picture had considerably brightened.

A TEST OF THE THEORY

To test my theory, I examined data from 41 *cantons* (counties) of Wallonia. The criterion for inclusion of a canton in this study is a minimum of 25 percent of its population engaged in agriculture in 1947. No Wallonian agricultural canton was excluded from the analysis. The Belgian canton unit best meets the criterion of systematic analyses that the unit be small enough to reduce internal variation of key variables, but large enough to constitute a whole.

Measures were found for the major variables. For economic activity, I measure the extent of market orientation in each unit by the mean price of a hectare of agricultural land.[8] I assume that the value of agricultural land measures the land's yield or productivity, and the greater the

---

[8] The mean price is figured by calculating the average value of arable and prairie land. Additionally, the data are from the 1895 agricultural census because data on land value at the cantonal level were not available for later periods. Since there is an interval of more than 40 years between the 1936 election and the 1895 census, this measure of land value lacks reliability. Though I have no data on this, I speculate that the relative differences in land value between cantons have not drastically changed between 1895 and 1936.

value, the more the unit's economic activity is capable of generating a surplus. The more this is the case, the more the region is characterized by a market orientation.[9] To measure property rights, I have examined the 1929 agricultural census, which provides three different measures: (1) the proportion of all medium-sized owner-cultivated farms; 2) the proportion of all small-sized tenant farms; 3) the proportion of all cultivators employed as agricultural laborers. To determine patterns of settlement, I have used data from the 1951 Belgian *Annexe Statistique* to measure mean cantonal population density.[10] The proportion of all votes for the Rexist party in the 1936 Belgian national legislative election is the measure of voting, the dependent variable.

I have also located measures for region and religiosity, two factors that emerged as strong predictors in my earlier study of French political regionalism. The effects of region and religiosity are tested against mode of production. To measure the effects of region, a dummy variable is created for each province by separating the cantons into those that actually belong to cantons in Luxembourg, Liége, Namur, Hainaut, and Brabant. To measure religiosity, I have used data from the 1952 *Carte de la pratique dominicale en Belgique par localités*. These data indicate on a scale of one to eight the degree to which a canton's population regularly attends Sunday church services.[11]

---

[9] The Belgian agricultural censuses limit the investigator's choice of reliable measures. Although land value is an adequate measure of land productivity, it may reflect, as well, other factors such as farm size, quality of soil, population density, and crop type.

[10] Due to the unavailability of data on population density for 1936, I employ the data from the 1951 Belgian *Annexe Statistique* to measure mean cantonal population density. Using the 1950 measure of population density, while less than optimal, is not totally inappropriate. Though I have no data on this, I speculate that population density generally varied little between 1936 and 1950.

[11] As is the case with population density there is no reliable measure of Belgian religiosity for 1936. I therefore employ the 1952 religiosity data on the reasonable assumption that religiosity did not vary greatly between 1936 and 1952. These data are the results of a study undertaken by the Centre Belge de Sociologie Religieuse. The Centre collected the data from the various parishes of the archbishoprics and bishoprics of Belgium. For the provinces of Brabant, Hainaut, and Liége, the raw data were absolute numbers of individuals regularly attending Sunday church service between 1950 and 1951 by parish. For the provinces of Luxembourg and Namur, the raw data were the results of a religious survey carried out by local clergy between 1946 and 1950. These data were estimations of the proportion of the population regularly attending Sunday church service by parish. The following are the percent equivalents for each of the eight intervals on the religiosity scale: 1 =

Table 1. Rexist Voting and Mode of Production Variables

Independent Variables	r
Land value	−.582*
Medium owner cultivation farms	.540*
Small tenant farms	−.637*
Agricultural laborers	−.467*
Population density	−.647*

* Significant at p less than .001.

Sources: *Statistique de la Belgique, Recensement Général de l'Agriculture de 1895.*

Statistique de la Belgique. Recensement Général de l'Agriculture de 1929.

Annexe Statistique, "Les résultats des élections prud'homales de Décembre 1950 avec la structure sociale des circonscriptions prud'homales."

Roger E. De Smet, René Evalenko, and William Fraeys, *Atlas des élections Belges 1919–1954* (Brussels: Université Libre de Bruxelles, 1958).

The first theory to be tested is that high Rexist voting corresponds to a mode of production characterized by a subsistence-oriented economy, medium-sized owner cultivation, and population dispersion. By contrast, low Rexist voting should correspond to a mode of production marked by a market-oriented economy, small scale tenancy or agricultural labor, and population concentration. To test this proposition, the proportion of Rexist voting is correlated with each of these mode of production indictors. Table 1 presents these correlations. The correlations demonstrate substantial support for the theory, and all are significant and in the predicted direction.

A combination of subsistence agriculture, medium-sized owner cultivation, and population dispersion should produce high Rexist voting, while a combination of a market-oriented agriculture, small-sized tenancy or agricultural labor, and population concentration should produce low Rexist voting. To test this theory, the five indicators of the mode of production are combined into a single index by summing the standardized scores for all measures. According to my theory, the variables of land value, small-sized tenant farms, agricultural laborers, and population density are all negatively related to the ideal pro-Rexist mode of production. To add them to the index, each is multiplied by (−1). In this way, all cases can be accurately arranged with respect to the ideal pro-Rexist

---

less than 20 percent who regularly attend Sunday church service, 2 = 20 to 25 percent, 3 = 25 to 40 percent, 4 = 40 to 50 percent, 5 = 50 to 60 percent, 6 = 60 to 75 percent, 7 = 75 to 82.5 percent, and 8 = greater than 82.5 percent. Also, the reader should be cautioned that these are cantonal level data and that I am predicting proportions for aggregates of farmers. Therefore, I am not committing the classical ecological fallacy because I am not predicting the behavior of individual cultivators.

mode of production. The proportion of Rexist voting is directly regressed on the index of the mode of production. Table 2 gives the results of the bivariate analysis of Rexist voting and mode of production.

The relationship is positive and significant at the .001 level, as expected. Nearly half of the explained variance in Rexist voting can be attributed to the mode of production.[12] The fit between mode of production and 1936 Rexist voting becomes clearer in Table 3. All 41 rural Wallonian cantons are listed in order of their mode of production index score. It is clear from this table that those cantons with positive index scores have the highest proportion of Rexist voting, while those with negative scores have the lowest proportion. When the political geography of Belgian Rexism is discussed, scholars point to Luxembourg province's strong support of the party and Hainaut province's feeble support. If my theory of voting is valid, Luxembourg's cantons should exhibit positive scores on the mode of production index and Hainaut's cantons negative scores. And, indeed, the data support my theory; 15 of the 16 Luxembourg cantons have a positive mode of production index score, while 5 of the 7 Hainaut cantons have a minus index score (see Table 3).

But skeptical readers may doubt that the mode-of-production theory is superior to other theories. One might, for instance, interpret the data as a consequence of some other unmentioned regional effect. If this were true, the relationship between voting and mode of production would be reduced to insignificance when controlled for region. When this is done, the relationship between mode of production and Rexist voting remains quite significant (see Table 4). When the effects of region are

---

[12] A fuller test of the explanatory power of the mode of production theory calls for an analysis of the 1939 Belgian legislative election as well as the 1936 election. In 1939, the Rexist party obtained only 4.4 percent of the national vote. However, the fascist character that might not have been totally apparent in 1936 had become quite clear by 1939. To test the strength of mode of production for the 1939 election, I regressed the proportion of Rexist voting in 1939 on the index of the mode of production for the same 41 Wallonian cantons. The explained variance is .528 for 1939 compared to .517 for 1936. This suggests that mode of production is a strong predictor of Rexist voting in both 1936 and 1939. Moreover, a fuller test of the theory calls for analysis of the effect of the percent rural. Though I limited the study to cantons with at least 25 percent farmers, there is substantial variation in the percent rural among the 41 cantons. To show that mode of production is a good predictor of Rexist voting when controlled for the percent rural, I regressed 1936 Rexist voting on mode of production and the percent rural. Mode of production emerged as the significant predictor with a standardized coefficient of .699, compared to .030 for the percent rural.

Table 2. Regression Analysis of Rexist Voting, Mode of Production, and Religiosity ($N = 41$)

	Independent Variables	
Dependent Variable	Mode of Production	Religiosity
Rexist voting		
Unstandardized	2.020*	
Standard error	.313	
Standardized	.719	
(Constant = 25.366)		
(R^2 = .517)		
Rexist voting		
Unstandardized	1.062*	3.305*
Standard error	.369	.871
Standardized	.378	.499
(Constant = 4.486)		
(R^2 = .650)		
Religiosity		
Unstandardized	.354*	
Standard Error	.057	
Standardized	.704	
(Constant = 5.207)		
(R^2 = .495)		

* Significant at p less than .001.

Sources: *Statistique de la Belgique, Recensement Général de l'Agriculture de 1895.*

Statistique de la Belgique, Recensement Général de l'Agriculture de 1929.

Annexe Statistique, "Les résultats des élections prud'homales de Décembre 1950 avec la structure sociale des circonscriptions prud'homales."

Roger E. De Smet, René Evalenko, and William Fraeys, *Atlas des elections Belges 1919–1954* (Brussels: Université Libre de Bruxelles, 1958).

Éditions du Dimanche-Mons. "Pratique Dominicale, d'après la Carte de la pratique dominicale en Belgique par localités." October 1952. Mons.

removed, the index for mode of production and Rexist voting remains significantly correlated.

Many scholars support a religiosity explanation of political regionalism (Derivry and Dogan 1986; Converse and Pierce 1986). They argue that distinct regional voting patterns can be attributed to regional variation in religiosity, as measured by church attendance. Frequent church attendance reflects a set of social values associated with a right-wing political ideology, while infrequent attendance reflects values associated with a left-wing political ideology. Since Rexism was an extreme right-wing political movement, areas of high religiosity should have exhibited strong support, while areas of low religiosity, weak support. As expected, the correlation coefficient between religiosity and Rexist voting is formidable, .683. To test the relative strength of mode of production and religiosity, I regressed Rexist voting on these two variables. Data in Table 2 show that mode of production and religiosity are both significant predictors of Rexist voting, although the standardized coefficient shows that the effect of religiosity is somewhat stronger. However, the two variables together explain

Table 3. Rexist Voting and Mode of Production

Canton	Province	Index Score	%Rexist Voting	%Agricultural
		Wallonian Cantons		
Fauvillers	Luxembourg	7	38	61
Sibret	Luxembourg	7	38	70
Houffalize	Luxembourg	5	37	54
Saint-Vith	Liége	5	24⁻	55
Erezée	Luxembourg	4	43	52
Larouche-en-Ardenne	Luxembourg	4	48	47
Neufchâteau	Luxembourg	4	41	45
Gedinne	Namur	4	39	46
Bastogne	Luxembourg	4	34	43
Saint-Hubert	Luxembourg	3	39	33
Wellin	Luxembourg	3	35	36
Nassogne	Luxembourg	3	20	31
Vielsalm	Luxembourg	3	31	34
Bouillon	Luxembourg	3	36	27
Etalle	Luxembourg	2	23	27
Paliseul	Luxembourg	2	35	33
Stavelot	Liége	2	24	40
Florenville	Luxembourg	2	18	29
Rochefort	Namur	1	31	29
Malmédy	Liége	1	31	35
Beaumont	Hainaut	1	15	38
Chimay	Hainaut	1	22	31
Beauraing	Namur	1	22	28
Philippeville	Namur	1	15	25
Ferrières	Liége	0	30	37
Durbuy	Luxembourg	0	35	35
Aubel	Liége	−3	26	30
Frasnes-les-Buissenal	Hainaut	−3	13	44
Ciney	Namur	−3	27	28
Lennik-St.-Quentin	Brabant	−4	6	37
Jodoigne	Brabant	−4	8	37
Perwez	Brabant	−4	16	29
Hannut	Liége	−4	18	35
Flobecq	Hainaut	−4	10	41
Eghezée	Namur	−5	26	31
Enghien	Hainaut	−5	8	30
Celles	Hainaut	−5	14	44
Chièvres	Hainaut	−5	7	31
Herve	Liége	−7	28	31
Jehay-Bodegnée	Liége	−7	17	25
Landen	Liége	−7	12	25

Sources: Statistique de la Belgique, Recensement Général de l'Agriculture de 1895.
Statistique de la Belgique, Recensement Général de l'Agriculture de 1929.
Statistique de la Belgique, Recensement de la Population de 1947.
Annexe Statistique. "Les résultants des élections prud'homales de Décembre 1950 avec la structure sociale des circonscriptions prud'homales." Roger E. De Smet, Réne Evalenko, and William Fraeys, *Atlas des élections Belges 1919–1954* (Brussels: Université Libre de Bruxelles, 1958).

more of the variance in Rexist voting than does either alone.

But before concluding that religiosity is a better predictor of Rexist voting than mode of production, I re-examined the relationship between religiosity and voting. First, advocates of the religiosity thesis do not explain why one set of social values should lead to right-wing voting and a different set to left-wing voting. While there may be a strong association between religiosity and voting, this may hide the operation of material interests. People who depend on the church for their material needs — education, employment, and charity — are likely

to support political programs that promote the welfare of their benefactor, the church. Since the Rexist party advocated programs favorable to the Catholic church, its supporters may have favored Rexism also for utilitarian reasons. Second, the association of voting and religiosity may be two effects of a common cause. That people hold conservative religious values does not explain why they vote for the political right. Their social values may merely reflect another side of their conservative political behavior. What must be explained is why some people have conservative religious values and others do not. In other words, the same structural factors

Table 4. Regression of Proportion Rexist Voting on Mode of Production and Region

Independent Variables	Unstandardized	Standardized
Mode of production	1.041* (.378)**	.370
Region[a]: Luxembourg	16.630* (4.571)**	.731
Region[b]: Liége	11.287* (4.455)**	.421
Region[c]: Namur	12.733* (4.874)**	.406
Region[d]: Hainaut	1.440* (4.571)**	.049
Constant	14.289	
R^2 .699		
$N = 41$		

Note: The four region variables are dummy variables that identify location in Luxembourg, Liége, Namur, and Hainaut provinces, respectively.

* Significant at p less than .001.

** Standard error for unstandardized coefficient.

Sources: Statistique de la Belgique, Recensement Général de l'Agriculture de 1895.

Statistique de la Belgique, Recensement Général de l'Agriculture de 1929.

Annexe Statistique. "Les résultants des élections prud'homales de Décembre 1950 avec la structure sociale des circonscriptions prud'homales."

Roger E. De Smet, Réne Evalenko, and William Fraeys, Atlas des Élections Belges 1919–1954 (Brussels: Université Libre de Bruxelles, 1958).

Éditions du Dimanche-Mons, "Pratique Dominicale, d'après la Carte de la pratique dominicale en Belgique par localités." October 1952. Mons.

that cause divergent political behavior may also be responsible for divergent religious behavior. Thus, the mode of production may be as good a predictor of religiosity as it is of voting. To test this theory, I regressed religiosity on the mode of production. The results in Table 2 show that mode of production is an excellent predictor of religiosity. The relationship is positive and significant at the .001 level, and mode of production accounts for nearly one-half of the explained variance in religiosity. This suggests that, rather than a cause of voting, religiosity may reflect another side of political behavior, and the mode-of-production model offers a valid explanation of both Belgian voting and religiosity in 1936. However, a strong correlation between mode of production and religiosity is no evidence of a causal explanation. The correlation may only indicate that a particular type of rural economic structure is associated with a certain level of religiosity.

CONCLUSION

The goal of this study was to add support to the mode-of-production theory of political behavior by showing that it applied to Belgium as well as France. The data strongly supported the theory

that divergent regional patterns of Belgian voting in 1936 could be largely ascribed to the existence of particular modes of production, which were responsible for distinctive regional constellations of interests. Because of its positions on the issues of large families, small- and medium-scaled private property, and corporatist society, Rexism received strongest support in regions characterized by a subsistence-oriented economy, medium-sized owner-cultivated farms, and population dispersion. It obtained its weakest support in regions marked by a market-oriented economy, small-sized tenant farms or farms employing agricultural wage labor, and population concentration. However, we must proceed cautiously in our assessment of the findings of this research. Because of measurement problems and my inability to incorporate some important contextual factors into my analysis, the conclusions drawn from these tests remain more suggestive that definitive. In future research, I hope to further demonstrate the validity of the mode-of-production theory to explain the political geography of German and Italian fascism.

REFERENCES

Annexe Statistique. 1951. "Les résultats des élections prud `homales de Décembre 1950 avec la structure sociale des circonscriptions prud `homales." vol. 1. Brussels: Ministry of the Interior.

Brustein, W. 1981. "A Regional Mode-of-Production Analysis of French Political Behavior: The Cases of Western and Mediterranean France." Politics and Society 10:355–98.

———. 1988. Social Origins of Political Regionalism: France, 1849–1981. Berkeley: University of California Press.

Cardoza, A.L. 1982. Agrarian Elites and Italian Fascism: The Province of Bologna, 1901–1926. Princeton: Princeton University Press.

Carpinelli, G. 1973. "Les Interprétations du Rexisme." Cahiers Marxistes (July–September): 71–82.

Chertok, R.H. 1975. "Belgian Fascism." Ph.D. Diss. Washington University.

Converse, P.E. and R. Pierce. 1986. Political Representation in France. Cambridge, MA.: Harvard University Press.

Corner, Paul. 1975. Fascism in Ferrara 1915–1925. London: Oxford University Press.

Daye, P. 1937. Léon Degrelle et le Rexisme. Paris: Fayard.

Degrelle, L. 1938. Révolution des Âmes. Paris: Fayard.

Denis, Jean 1936. Principes Rexistes. Brussels: Rex.

Derivry, D. and M. Dogan. 1986. "Religion, Classe et Politique en France." Revue Française de Science Politique. 36:157–80.

Éditions du Dimanche-Mons. 1952. "Pratique Dominicale, d'après la Carte de la pratique dominicale en Belgique par localités." October: Mons.

Eley, G. 1983. "What Produces Fascism: Preindustrial Traditions or a Crisis of the Capitalist State." Politics and Society 12:53–82.

Étienne, J.M. 1968. *Le Mouvement Rexiste jusqu'en 1940*. Paris: Colin.

Faris, E., 1975. "Takeoff Point for the Nationalist Socialist Party: The Landtag Election in Baden, 1929." *Central European History*, 8:140–71.

de Felice, R. 1966. *Mussolini il Fascista*, vol. 1. Torino: Giulio Einaudi.

Fitzmaurice, J. 1983. *The Politics of Belgium*. London: Hurst & Co.

Fromm, E. 1966. *Escape from Freedom*. New York: Avon Books.

Goris, J-A. 1946. "Belgium Between the Two World Wars." Pp. 66–80 in *Belgium*, edited by J-A. Goris. Berkeley: University of California Press.

Hamilton, R.F. 1982. *Who Voted For Hitler?* Princeton: Princeton University Press.

Henry, A. 1937. *Les Variations Régionales de l'Agriculture en Belgique*. Brussels: Secrétariat de la Société Scientifique de Bruxelles.

Hojer, C-H. 1946. *Le Régime Parlementaire Belge de 1918 à 1940*. Uppsala, Sweden: Almqvist & Wiksells.

Kornhauser, W. 1959. *The Politics of Mass Society*. Glencoe, IL: Free Press.

Lenski, G. and J. Lenski 1982. *Human Societies*, 4th ed. New York: McGraw-Hill.

Lesthaeghe, R.J. 1977. *The Decline of Belgian Fertility, 1800–1970*. Princeton: Princeton University Press.

Linz, J. 1976. "Some Notes Toward a Comparative Study of Fascism in Sociological Historical Perspective." Pp. 3–121 in *Fascism: A Reader's Guide*, edited by W. Laquer. Berkeley: University of California Press.

Lipset, S.M. 1981. *Political Man: The Social Bases of Politics*. Baltimore: Johns Hopkins University Press.

Marx, K. [1859] 1959. *Marx and Engels: Basic Writings on Politics and Philosophy*. Lewis Feuer (ed.) Garden City, NY: Anchor Books.

Misner, E.G. 1946. "The Agriculture of Belgium." Pp. 148–66 in *Belgium*, edited by J.A. Goris. Berkeley: University of California Press.

Moore, Barrington Jr. 1966. *Social Origins of Dictatorship and Democracy*. Boston: Beacon Press.

Nolte, E. 1966. *Three Faces of Fascism*. New York: Holt, Rinehart, and Winston.

Palme Dutt, R. 1974. *Fascism and Social Revolution*. San Fancisco: Proletarian Publishers.

Parsons, T. 1954. *Essays in Sociological Theory*. Glencoe, IL: Free Press.

Passchier, N. 1980. "The Electoral Geography of the Nazi Landslide." Pp. 283–300 in *Who Were the Fascists*, edited by Larsen, Hagtvet, and Myklebust. Bergen, Norway: Universitetsforlaget.

Petersen, J. 1975. "Elettorato e Base Sociale del Fascismo Italiana negli Anni Venti." *Studi Storici* 16:627–69.

Pollock, J.K. 1944. "An Areal Study of the German Electorate, 1930–1933." *American Political Science Review*. 38:89–95.

Poulantzas, Nicos. 1974. *Fascism and Dictatorship*. London: New Left Books.

Pridham, G. 1973. *Hitler's Rise to Power: The Nazi Movement in Bavaria, 1923–1933*. New York: Harper & Row.

Quicke, F. 1966. *Les Régions Agro-Économiques de la Belgique*. Liège: Sciences et Lettres.

Reich, W. 1972. *The Mass Psychology of Fascism*. London: Souvenir Press.

Schepens, Luc 1980. "Fascists and Nationalists in Belgium 1919–1940." Pp. 501–16 in *Who Were the Fascists*, edited by Larsen, Hagtvet, and Myklebust. Bergen, Norway: Universitetsforlaget.

de Smet, R.E., R. Evalenko, and W. Fraeys. 1958. *Atlas des Élections Belges 1919–1954*. Brussels: Université Libre de Bruxelles.

Snowden, F.M. 1972. "The Origins of Agrarian Fascism in Italy." *Archives Européenes de Sociologie*. 13:268–95.

Statistique de la Belgique. 1899. *Recensement Général de l'Agriculture de 1895*, vol. 3, Brussels: Ministry of Agriculture and Public Works.

Statistique de la Belgique. 1937. *Recensement Général de l'Agriculture de 1929*, Brussels: Ministry of the Interior (Central Office of Statistics).

Statistique de la Belgique. 1949. *Recensement de la Population de 1947*, Brussels: Ministry of Economic Affairs (National Institute of Statistics).

Stengers, J. 1965. "Belgium." Pp. 128–67 in *The European Right. A Historical Profile*, edited by Rogger and Weber. Berkeley: University of California Press.

Wallef, D. 1980. "The Composition of Christus Rex." Pp. 517–23 in *Who Were the Fascists*, edited by Larsen, Hagtvet, and Myklebust. Bergen, Norway: Universitetsforlaget.

Weber, E. 1964. *Varieties of Fascism*. Princeton: D. Van Nostand.

THE GENERALITY OF DEVIANCE
IN LATE ADOLESCENCE AND EARLY ADULTHOOD*

D. WAYNE OSGOOD
University of Nebraska, Lincoln

LLOYD D. JOHNSTON,
PATRICK M. O'MALLEY,
JERALD G. BACHMAN
The University of Michigan

Because a wide variety of deviant behaviors are positively correlated with one another, some researchers conclude that all are manifestations of a single general tendency. The present analysis incorporated three waves of self-reports about heavy alcohol use, marijuana use, use of other illicit drugs, dangerous driving, and other criminal behavior for a nationally representative sample of high school seniors. A relatively stable general involvement in deviance accounted for virtually all association between different types of deviance, but the stability of each behavior could only be explained by equally important and stable specific influences. Thus, theories that treat different deviant behaviors as alternative manifestations of a single general tendency can account for some, but far from all, of the meaningful variance in these behaviors. The only significant influence of one type of deviance on another was that of marijuana use on later use of other illicit drugs. The causal model also revealed interpretable shifts in the associations among these behaviors over the four years following high school.

INTRODUCTION

Research has firmly established that a wide range of deviant behaviors are positively correlated with one another during adolescence and early adulthood (e.g., Akers 1984; Donovan and Jessor 1985; Elliott and Huizinga 1984; Johnston, O'Malley, and Eveland 1978). This paper concerns the sources of that association and its significance for theories of deviance. We will be particularly concerned with the possibility that deviance is a unified phenomenon, with various behaviors serving as alternative manifestations of a more general tendency.

There are two plausible general explanations for correlations among deviant behaviors. The first is that engaging in one form of deviant behavior leads to engaging in others as well. Many people believe that there are causal links between some forms of deviance, particularly that drug use leads to crime. The second explanation is that different deviant behaviors are related because they have shared influences. For example, the factors that lead people to become sexually active at an early age might be the same as (or at least overlap) those that lead them to use marijuana. To the degree that the same factors are major sources of all deviant

behaviors, it is meaningful to speak of a general syndrome of deviance (Donovan and Jessor 1985).

Sociologists have offered many definitions of deviance. (For a general discussion, see Gibbs 1981, ch. 2.) Our concerns center on behavior socially defined as undesirable rather than on the social processes that lead certain individuals to be labeled deviants. Jessor and Jessor (1977) offer a clear definition of deviance (which they refer to as problem behavior): it is "behavior that is socially defined as a problem, a source of concern, or as undesirable by the norms of conventional society and the institutions of adult authority, and its occurrence usually elicits some kind of social control response." (p. 33) Our study examines several deviant behaviors.

By definition, all deviant behaviors violate conventional standards of behavior. Even so, each deviant behavior may be a unique phenomenon requiring a separate explanation, or the various deviant behaviors may form a unified phenomenon with a single explanation. This is an empirical question with major import to theories of deviance. The generality of deviance across different types of behavior will be a function of the degree to which the behaviors have the same influences.

. Generality versus specificity is a relevant issue in many areas of sociology. For instance, we speak of social status as encompassing income, education, and occupational prestige; the transiency, poverty, and physical deterioration of a community are all signs of social disorganization; and social stratification includes income inequality, residential segrega-

tion, and social mobility. These are meaningful groupings that have conceptual coherence, but they may or may not constitute unified empirical phenomena. If social stratification is empirically general across its different manifestations, then explaining one form of stratification is sufficient to explain them all. In this case, plausible theories of stratification would look quite different than if the different forms occur independently of one another. Our framework for assessing the generality of deviance is applicable to other concepts as well.

Our approach to analyzing generality versus specificity transcends any particular theoretical position. We do not focus on the role of an a priori set of explanatory variables, such as Jackson et al.'s (1986) assessment of specificity based on differential association theory. Instead, we partition all reliable variance into general and specific components on the basis of covariance among various deviant behaviors in a longitudinal research design. ·

SHARED INFLUENCES AND THE GENERALITY OF DEVIANCE

Most sociological theories are consistent with deviance being general across different behaviors. Almost any explanation offered for one behavior has been offered for others as well. For instance, social scientists have argued that peer influence leads to early cigarette smoking (Krosnick and Judd 1982), early sexual intercourse (Billy and Udry 1985), marijuana use (Kandel 1978), and criminal behavior (Sutherland and Cressey 1955). Models emphasizing social learning (Akers 1977), subcultural norms (Coleman 1961), enhancement of self-esteem (Kaplan 1975), and social bonds (Hirschi 1969) have been proposed for a variety of deviant behaviors, and some theorists have simultaneously addressed several forms of deviance within a single explanatory framework (e.g., Elliott, Huizinga, and Ageton 1985; Jessor and Jessor 1977; and Kaplan 1975).

While many social explanations indicate how different deviant behaviors might have influences in common, explanations vary in the degree to which the process causing one deviant behavior will jointly produce others. For instance, Akers' social learning approach (1977) explains each deviant behavior as a result of associating with people who model and reinforce that behavior. The processes of modeling and reinforcement lead to both alcohol use and theft only to the degree that associates who support one also support the other. Thus, in Akers' theory the generality of deviance is dependent on the empirical correlation between conceptually independent variables.

The strongest position in favor of the generality of deviance is that different deviant behaviors are manifestations of a single underlying construct. Jessor and his colleagues have posited that a variety of deviant behaviors form a "syndrome," which is directly caused by a general latent variable of unconventionality (Donovan and Jessor 1985; Jessor and Jessor 1977). Hirschi (1984) explains the relationship between drug use and delinquency in a similar manner, stating that the two are not merely influenced by some of the same factors, but "they are manifestations of the same thing" (p. 51). This "thing" is criminality, which he defines as "the tendency or propensity of the individual to seek short term, immediate pleasure" (p. 51). According to Hirschi's social control theory (1969), criminality results from the absence of social bonds. Hirschi and Gottfredson have recently articulated wide-ranging theoretical ramifications of the concept of criminality, with its image of deviant behavior as a manifestation of general and relatively stable individual differences (Gottfredson and Hirschi 1986; Hirschi and Gottfredson 1983, 1986). For our purposes, the most important implication of the positions taken by Jessor and colleagues and by Hirschi and Gottfredson is that explaining a general tendency toward deviance is sufficient to account for a large group of behaviors and that causes specific to any particular form of deviance are relatively unimportant.

The intermediate position is that a general cause, alienation from the norms of conventional society, is a *partial* determinant of a range of deviant behaviors. Cloward and Ohlin (1960) argue that the community's illegitimate opportunity structure shapes the specific deviance that will result, while Elliott et al. (1985) point to social learning from the peer group. Such theories predict substantial, but not complete, generality of deviance.

INFLUENCE OF ONE DEVIANT BEHAVIOR ON ANOTHER

Taking the notion of the generality of deviance to its limit would preclude any influence of one type of deviant behavior on another. In this case, influences specific to particular forms of deviance not only would be unimportant, but nonexistent. Given a propensity toward deviance, the specific deviant behaviors in which a person engages at any time would be strictly random. It then follows that there would be perfect correlations among different deviant behaviors, limited only by the reliability of their measures. Since there would be nothing unique about specific behaviors, they could not possibly influence one another.

Short of this extreme position, however, it is

possible for different deviant behaviors to have some influences in common and some influences that are specific, including one behavior serving as a partial cause of another.

Though we know of no theories that explicitly predict influences between specific deviant behaviors, such influences would be consistent with several theories.[1] Consider a person who begins to use marijuana regularly. This behavior could lead to rejection by conventional peer groups and increased association with peer groups that approve of various forms of deviance. If so, social learning from the new group could result in other deviant behaviors, as could weakened bonds to conventional groups. If the marijuana use was detected by authority figures, then labeling theory predicts the development of a deviant identity and secondary deviance (Lemert 1972), which presumably would encompass a variety of behaviors.

Available Evidence

Several studies are pertinent to the generality of deviance and the sources of positive correlations between various deviant behaviors. Findings about shared influences appear in studies that relate the same explanatory variables to several different deviant behaviors and in analyses of the factor structure of covariance among behaviors. Influence between specific deviant behaviors has been investigated through longitudinal research measuring the same behaviors at two or more times.

Elliott et al. (1985) and Jessor and Jessor (1977) investigated a variety of causal factors, and their findings support the possibility that shared influences create relationships between different deviant behaviors. Factors that strongly influenced one deviant behavior (such as delinquency) similarly influenced other behaviors (such as alcohol and drug use).

On the other hand, some research indicates that certain causal factors are more important for one deviant behavior than for others. For instance, Kandel, Kessler, and Margulies (1978) concluded that parental influences were much less important for marijuana use than for use of other illicit drugs. Johnston (1973) found that ideological alienation related to some forms of illicit drug use, but not others, and not to cigarette use, alcohol use, or delinquency. These findings imply that involvement in deviance is not completely general across behaviors.

Studies such as these have limited value for

determining the importance of shared versus specific influences on different deviant behaviors. Whatever the findings, their implications are limited to the role of the finite set of explanatory variables included in the research. The shared influences identified by Elliott et al. and by Jessor and Jessor might be of little consequence compared to specific influences from variables they did not study. Conversely, the specific influences identified by Kandel et al. and Johnston might be trivial departures from a larger pattern of shared influences. Directly assessing the generality of deviance requires an approach that is not limited to measured influences.

The work of Donovan and Jessor (1985) illustrates such an approach. They focused on covariance among deviant behaviors rather than on the relationships of the behaviors to potential explanatory variables. Using confirmatory factor analysis, Donovan and Jessor determined that a variety of behaviors formed a general syndrome of deviance. Their results indicate that a single latent variable is sufficient to account for covariance among the behaviors, and they replicate this result for several samples.

Nevertheless, further evidence is needed before accepting Donovan and Jessor's conclusions. Their analyses are cross-sectional, and, therefore, they could not distinguish between covariance due to shared influences and covariance due to an influence of one behavior on another. Furthermore, not all behaviors were well explained by the general syndrome. In many instances, the latent variable accounted for less than 10 percent of the variance of specific behaviors. Though their conclusions imply that any remaining variance was simply error of measurement, their method provides no means of differentiating error of measurement from meaningful variance specific to a particular variable.

Two studies have used longitudinal data from nationally representative samples to assess influence between crime (or delinquency) and illicit drug use. Both Johnston, O'Malley, and Eveland (1978) and Elliott and Huizinga (1984) concluded that shared influences are the major source of the relationship between these behaviors and that influence of one behavior on another is relatively unimportant. Nevertheless, neither study strictly rules out the possibility of influence between behaviors, and each provides some evidence of such influences. Johnston et al.'s cross-lag panel analysis yielded path coefficients consistent with modest reciprocal influence between the behaviors (p. 151). Similarly, Elliott and Huizinga found that earlier delinquency predicted later drug use, even after taking earlier drug use into account (p. 88).

The present study uses modern techniques of

---

[1] The possibility that physical need due to drug addiction leads to increased crime falls outside our interest in social explanations of deviance, and physical addiction is rare in our sample.
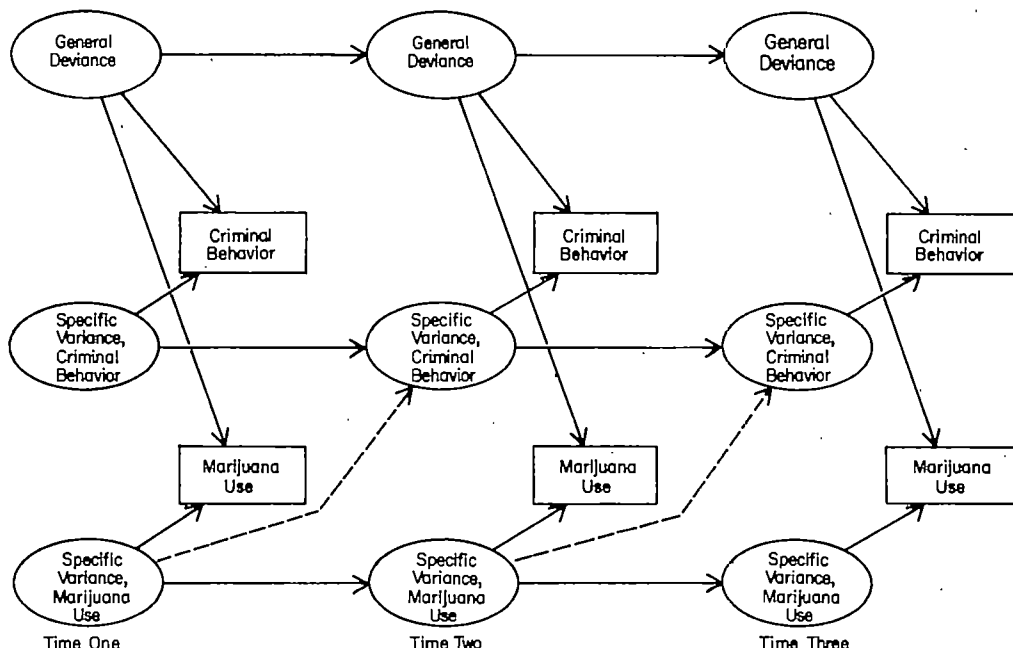
Fig. 1. A Causal Model Differentiating General and Specific Influences on Deviant Behaviors.

causal modeling for panel data (e.g., Kessler and Greenberg 1981) to advance our understanding of the relationships between different forms of deviance. We develop a structural equation model that separates general and specific components of each behavior. The longitudinal aspect of our study allows us to examine the importance of general involvement in deviance not only in terms of the size of the general component for each behavior (as in Donovan and Jessor's 1985 analysis), but also to compare it directly to reliable specific variance and to assess the stability of both over time. Our test for influence between specific variables improves on earlier work by addressing several types of deviance, by explicitly modeling shared influences, and by allowing for error of measurement. Finally, by covering several years of the age span, the model can relate changing age norms to shifts in the strength of connections between individual behaviors and a general syndrome of deviance.

METHOD

Causal Models

Our basic causal model is illustrated in Figure 1. For simplicity of presentation, this figure is limited to two behaviors: criminal behavior and marijuana use. The model incorporates a general tendency toward deviance as a contemporaneous influence on all of the behaviors. It allows for longitudinal influences in the form of stability of that general tendency, stability of variance

specific to each behavior, and influence of one behavior on another.

The causal model is unusual in that each behavior serves as an indicator of the general factor and also has unique aspects that may influence other variables. In effect, we divide measures of behavior into three components: variance shared with a general tendency toward deviance; reliable variance specific to the behavior; and error of measurement. The model includes disturbance terms for all latent variables as well.

The model will be identified given three conditions. First, there can be no correlated errors between the general and specific latent variables. Second, there must be constraints on the error terms of the observed variables, or the specific variances will be undefined. We obtained meaningful estimates by assuming that all influence of wave one on wave three is mediated by wave two and that reliability is constant over time[2] (Kessler and Greenberg

[2] An equally plausible assumption is that error variance, rather than reliability, was constant. For these data, the assumption of equal reliability proved superior; it yielded consistent and interpretable estimates, while the assumption of constant error variance resulted in negative estimates of residual variance. Because reliability refers to the proportion of error variance (rather than the absolute amount of error variance), it was necessary to estimate the model from the correlation matrix. The presented coefficients have been transformed to reflect the original metric of the variables (as in an analysis of a covariance matrix).

1981, p. 147–50). Third, the number of paths among the specific variances must be limited. Within each wave, one degree of freedom per behavior is absorbed by estimating influences from general deviance, and one degree of freedom between adjacent waves is absorbed by estimating the stability of general deviance. The former reduces the number of possible error paths among contemporaneous specific variances, while the later reduces the number of possible longitudinal influences between them.

The importance of the general tendency is reflected both by the strength of its paths to each behavior and by its stability over time. If all the behaviors are interchangeable manifestations of the general tendency (the extreme version of the shared-influence hypothesis), there will be no meaningful variance specific to the separate behaviors. In this case, a simpler model would be adequate to account for the data, a model in which each behavior is comprised only of general deviance and error of measurement, and the stability of general deviance is the only longitudinal influence. On the other hand, the presence of a substantial amount of stable specific variance would indicate that the various forms of deviance cannot be fully explained by a single causal process.

Because the model in Figure 1 divides each deviant behavior into general and specific components, it separates the influence of one behavior on another from their associations with general deviance. Influences between behaviors are indicated by paths from one specific variance to a different specific variance at the next time. The dashed lines in Figure 1 represent such an influence of marijuana use on criminal behavior.

It is important that the model also take into account error of measurement; failing to do so would cause bias in estimates of causal paths. Given a set of uniformly positively related variables such as these, the bias will be toward spurious positive influences among behaviors. This might explain the apparent influence between drug use and delinquency in earlier research (Elliott and Huizinga 1984; Johnston et al. 1978).

Sample

The data we analyze were collected as part of the Monitoring the Future study (Osgood, Johnston, O'Malley, and Bachman 1985). For a detailed description of the sample design and data collection, see Bachman and Johnston (1978). For a full listing of variables, see Johnston, Bachman, and O'Malley (1986). For detailed findings on drug use and related variables, see Johnston, O'Malley, and Bachman (1986). In this ongoing study, which began

in 1975, a wide range of information is gathered from a nationally representative sample of high school seniors each year.[3] Data for the present analysis come from the follow-up portion of the study, which is based on a subsample of each senior class. Half of the participants in the follow-up study complete questionnaires in every odd-numbered year after graduation, and the other half do so in every even-numbered year. The present analysis used three waves of data, provided at approximate ages of 18, 19, and 21, or 18, 20, and 22. The analysis was limited to white respondents, since the black subsample is somewhat less representative due to differential high school drop-out rates. The follow-up study over-samples the more serious drug users in high school to obtain more accurate estimates for this segment of the population. The over-sampled individuals are then given smaller weights in analyses to produce a representative sample. There were 975 respondents in the sample, coming from the high school senior classes of 1976 though 1980, yielding a weighted sample of 717 cases.

Initial involvement in the behaviors being studied typically occurs at earlier ages than are included in this sample (e.g., Elliott and Huizinga 1984). We do not consider this a shortcoming of our study. We reject the point of view that a deviant behavior is "caused" at some time of "onset," after which it is self-perpetuating until it is "caused" to stop. Instead, we assume that a behavior occurs when its causes are present and does not occur when they are absent. As our results demonstrate, the stability of these behaviors is far from perfect, meaning that there is a great deal of change independent from the general age trends. Thus, there is just as much need to explain persistence of deviant behavior as to explain its onset and cessation. As Hirschi (1984, p. 50) has articulated, it is not at all clear that "onset" is a meaningful concept for deviant behavior. It is only in retrospect that getting drunk for the first time can be called the onset of alcoholism. Asking whether the first marijuana cigarette precedes the first incident of theft is a much less meaningful question than asking whether current marijuana use has an influence on later criminal behavior.

---

[3] Each year, a three-stage national probability sample leads to questionnaire administrations in approximately 130 high schools (roughly 110 public and 20 private). This procedure yields between 15,000 and 19,000 respondents. A random one-fifth of each annual sample completes the version of the questionnaire that includes the items used in the present analysis. Response rates for the base year average 80 percent, and follow-up response rates are generally 75 percent or more of the original group.

Table 1. Goodness of Fit of Alternative Models

		Δ	d.f.	χ^2	p
I.	Null Model	.000	105	4,563.94	0.
Measurement models					
II.	All possible within-wave relationships, no longitudinal relationships	.461	75	2,461.63	0.
III.	One general factor at each wave, no correlated errors, no longitudinal relationships	.427	90	2,615.21	0.
IV.	One general factor at each wave, 11 correlated errors, no longitudinal relationships	.460	79	2,465.94	0.
Longitudinal models					
V.	Stable general factor, no stable specific factors	.715	78	1,298.88	0.
VI.	Stable general and specific factors	.982	63	80.11	.07
VII.	Stable general and specific factors, plus longitudinal influence among all specific factors	.992	43	36.07	.76
VIII.	Stable general and specific factors, plus influence of marijuana use on other illicit drug use	.986	61	63.90	.38

Note: Models V–VIII include 10 correlated errors (see footnote 5).

Measures

Our analysis is based on self-report measures of five different types of deviant behavior: criminal behavior (limited to illegal behavior directed at victims); heavy alcohol use; marijuana use; use of other illicit drugs; and dangerous driving. We chose these behaviors because they represent a broad range of the conventionally proscribed activities that are common during this age span. Research on the factor structure of substance abuse has shown that use of alcohol, marijuana, and hard drugs are relatively distinct phenomena (e.g., Hays et al 1986), so we consider it appropriate to treat them separately. Dangerous driving is not among the deviant behaviors typically studied by social scientists. Nevertheless, it is quite appropriate to our definition of deviance since it is generally recognized as undesirable and is subject to social controls. Furthermore, it is an exciting, risky activity that follows the same age trend as criminal behavior (Hirschi and Gottfredson 1983). These five behaviors do not exhaust the concept of deviance, and our results may or may not characterize other types of deviance.

The measure of heavy alcohol use referred to behavior in the past two weeks, and the other four measures referred to behavior over the last 12 months. The 14-item measure of criminal behavior also was used in the Youth in Transition study (Bachman, O'Malley, and Johnston 1978), and is an adaptation of Gold's (1970) well-known measure. These items concern interpersonal aggression, theft, and vandalism.[4] Our index was a sum across the items,

each of which ranged from zero to four (for committing the act five or more times). Heavy alcohol use was measured in terms of the number of occasions a respondent had five or more drinks in a row during the last two weeks, with scores ranging from zero through five (10 or more). The scale for marijuana use varied from zero through nine (40 or more times in the last month). Use of other illicit drugs was measured as an average across eight drugs, each scored on the same scale as marijuana use. The measure of dangerous driving was the sum of reports of traffic tickets and traffic accidents, each scored as zero through four (four or more).

RESULTS

Goodness of Fit of Alternative Models

We compared the goodness of fit of several alternative causal models to determine which explanatory factors account for observed relationships between the five deviant behaviors. The factors of interest were a general tendency toward deviance (representing shared influences); variance specific to a particular behavior (representing specific influences); and influences between particular behaviors. Table 1 summarizes the fit of the models, which were estimated by LISREL IV (Joreskog and Sorbom 1978). Chi-square values and their associated

---

[4] The label "criminal behavior" is simply heuristic, since the other types of deviant behavior also may involve violating laws. The label should not be too misleading, however, since this aggregation is limited to

actions that intentionally victimize other people. The other measures pertain to either victimless crimes or traffic offenses (where victimization is rarely intended). For respondents who had not yet reached the age of majority, the variable technically refers to delinquency rather than crime. Some analyses of these items have distinguished interpersonal aggression from property offenses (e.g., Bachman et al. 1978). Preliminary analyses revealed negligible reliable variance in aggressive offenses at waves 2 and 3, meaning that separating these offenses would be fruitless for the present study.

probability levels indicate the lack of fit between the models and the observed covariances among the 15 variables (five deviant behaviors, each measured on three occasions). The probability levels should not be taken literally, however, since these skewed data do not justify the assumption of multivariate normality. A particularly useful index of fit is Bentler and Bonett's Δ (1980), which is independent of sample size. Model I is the "null model," which corresponds to the assertion that all of the variables are unrelated, and Δ is the proportionate reduction of this chi-square provided by other models.

Measurement model. Our strategy in developing the causal model was to obtain an adequate within-wave measurement model before testing alternative longitudinal models. Our goal for the measurement model was to divide the reliable variance of each behavior into general and specific components. Cross-sectional covariance among the five behaviors provides a basis for defining a general factor of deviant behavior. Donovan and Jessor (1985) concluded that such a single latent variable was sufficient to account for cross-sectional relationships among a variety of deviant behaviors.

Model II serves as a standard for the fit of the measurement model to the cross-sectional relationships, since it incorporated all possible relationships within each wave but did not allow any relationships between waves. The reduction in χ^2 provided by this model is equal to the sum of the χ^2 values for the null models of the three within-wave covariance matrices. For Model II, $\Delta = .461$, meaning that 46 percent of the total chi-square value was due to cross-sectional relationships and 54 percent was due to longitudinal relationships.

Model III is the basic measurement model, allowing for a single general factor at each wave. This model explained a large share of the cross-sectional relationships ($\Delta = .427$), though a significant amount of within-wave covariation remained (comparing Models II and III: $\Delta_{\text{II-III}} = .034$, d.f. = 15, $\chi^2 = 153.58$, $p = 0.$). This finding is in substantial agreement with the results of Donovan and Jessor (1985), in that a general factor accounts for 93 percent of the chi-square value attributable to within-wave relationships ($\Delta_{\text{III}}/\Delta_{\text{II}}$). Though we found significant lack of fit for a single-factor model where they did not, this is likely due only to our larger sample size.

The discrepancy in fit between Models II and III shows that some pairs of deviant behaviors were more strongly correlated with one another than is consistent with a single-factor model. By allowing some correlated errors between behaviors, we obtained more accurate estimates of the variances shared with the general factor. Furthermore, failing to incorporate such paths in other models could lead to spurious longitudinal paths.

Model IV allowed for several correlated errors within each wave in addition to a general factor. This model accounted for virtually all of the within-wave covariance ($\Delta_{\text{II-IV}} = .001$, d.f. = 4, $\chi^2 = 3.31$, $p > .25$), and it served as the measurement model for the longitudinal models. It incorporated correlated errors between heavy drinking and marijuana use, marijuana use and other illicit drug use, criminal behavior and dangerous driving, and heavy drinking and dangerous driving.[5]

Longitudinal models. Model V was estimated to test the extreme version of the shared-influences hypothesis. In this model, all longitudinal relationships are explained by the stability of a general tendency toward deviance. The only latent variable at each wave was general deviance, so all specific variance was treated as error variance in the observed variables.

Model V accounted for roughly half of the longitudinal covariation ($\Delta_\text{V} = .715$; $\Delta_{\text{V-II}}/[1 - \Delta_{\text{II}}] = .47$), and left a highly significant and substantively important portion of the total covariance unexplained. Clearly, an adequate model requires longitudinal influences involving specific components.

The remaining models are variations on the model illustrated in Figure 1. Each observed variable was modeled as a function of a latent variable of general deviance, a latent variable of specific variance, and error of measurement. These models included paths for the stability of the latent variables, and we assumed that all influence of wave one on wave three was mediated by wave two and that reliability was constant over time.

Model VI expanded on Model V by allowing for a stable, specific component of each behavior as well as for general deviance; it did not allow for any influences between different forms of deviance. This model fit the data quite well ($\Delta = .982$), indicating that the vast majority of longitudinal covariation was attributable to the stabilities of the general factor and of the reliable variance specific to each behavior.

---

[5] To avoid inflated estimates of the variance each measure shared with the common factor, only positive correlated error paths were included. The path between heavy drinking and dangerous driving became negative and insignificant for the third wave, so it was eliminated. In models incorporating longitudinal influences (Models V—VIII), the path between marijuana use and other illicit drug use was removed for the same reason. The correlated error paths were between the latent variables in Models VI—VIII and between the observed variables in Models IV and V (which did not have latent variables for specific variance).

Table 2. Variance Components and Reliabilities

	Mean	Variance	Rel.	Variance Components			Proportion of Reliable Var.	
				General	Specific	Error	General	Specific
Criminal behavior								
Time 1	2.77	16.59	.70	8.68	3.12	4.99	74%	26%
Time 2	1.98	12.06	.70	4.60	3.81	3.63	55%	45%
Time 3	1.37	7.13	.70	2.47	2.51	2.15	50%	50%
Heavy alcohol use								
Time 1	.97	1.80	.70	.57	.69	.54	45%	55%
Time 2	1.09	1.86	.70	.49	.80	.55	38%	62%
Time 3	1.08	1.89	.70	.40	.90	.56	31%	69%
Marijuana use								
Time 1	2.06	7.99	.90	2.47	4.47	.82	36%	64%
Time 2	2.29	8.17	.90	2.52	4.60	.84	35%	65%
Time 3	2.16	7.99	.90	2.52	4.58	.82	35%	65%
Other illicit drug use								
Time 1	.12	.139	.76	.040	.064	.034	38%	62%
Time 2	.14	.135	.76	.044	.059	.033	43%	57%
Time 3	.16	.159	.76	.057	.066	.039	46%	54%
Dangerous driving								
Time 1	.81	1.58	.49	.21	.56	.81	27%	73%
Time 2	.83	1.38	.49	.21	.46	.71	32%	68%
Time 3	.64	1.04	.49	.14	.36	.53	29%	71%

We evaluated the fit of Model VII as a general test of the hypothesis that there are longitudinal influences of some specific deviant behaviors on others. This model allowed for influence of each specific factor on all others at the subsequent wave. For this model to be identified, it was necessary to constrain between-behavior influences from wave one to wave two to be equal to those from wave two to wave three.[6] This constraint also increases the power of the test, provided influences are roughly similar across the two time intervals. Model VII yielded a small, but significant, improvement in fit, which indicates the presence of influence between specific deviant behaviors (Δ_{VII-VI} = .010, d.f. = 20, χ^2 = 44.01, p = .001).

The coefficients for Model VII and the residuals and first derivatives for Model VI suggested that the strongest influence between specific factors was that of marijuana use on later use of other illicit drugs. In Model VIII, this path was added to those allowed in Model VI, yielding Δ = .986 and significant reduction in chi-square ($\Delta_{VIII-VI}$ = .004, d.f. = 2, χ^2 =

16.21, p < .001). The fit of this model was excellent, with the chi-square value virtually equal to the number of degrees of freedom. Furthermore, adding other paths between specific deviant behaviors did not significantly improve the fit of this model. Thus, we found evidence of influence between specific forms of deviant behavior, but this influence was very circumscribed.[7]

Path Estimates

Variance components. Table 2 shows the means and variances of the five measures for each wave, along with the division of the variance into general, specific, and error components that is implied by the path estimates of Model VIII.[8] Our presentation of the model departs from common practice by emphasizing variance components and explained variance as much as

---

[6] Identification becomes an issue because there are only as many degrees of freedom arising from correlations between adjacent waves as there are possible paths between specific behaviors. Allowing all of these paths would leave no degrees of freedom for the stability of general deviance. While our solution of assuming equal influence across both intervals provides sufficient constraints to generate estimates, it is evident from the very large standard errors for all longitudinal paths that the model is empirically under-identified. Thus, Model VII is useful for assessing the potential improvement in fit from cross-behavior influences, but it does not provide meaningful estimates of those influences.

[7] Though this is only one significant cross-behavior path out of a possible twenty, we are confident that it is not a chance relationship. Comparing Models VI and VII gives clear evidence of influence between specific behaviors, and the pair of paths from marijuana use to later use of other illicit drugs is significant far beyond the chance level of .05. Even allowing for the non-normality of our data, it is not plausible that a relationship of this magnitude would occur by chance.

[8] The variance components are equal to the square of a measure's loading on the relevant latent variable (lambda) times the variance of that latent variable. The variance components do not sum to the exact amount of total variance because LISREL is not constrained to precisely reproduce the diagonal of the covariance matrix. The metric of general deviance was set by fixing the lambdas for marijuana use to one.

or more than path estimates. While this would not be desirable for most models, it is quite useful given our interest in comparing the importance of general and specific explanations for each deviant behavior. The variance components provide a straightforward comparison by combining the variance of the latent variables with their loadings on the observed variables, placing general and specific components in a shared metric.

Means and variances were relatively constant over time for heavy drinking, marijuana use, and other illicit drug use, but both statistics declined for criminal behavior and dangerous driving. This age trend is well documented for both of these behaviors (Hirschi and Gottfredson 1983).

Reliability is defined as the proportion of non-error variance for a measure. The measure of marijuana use had the highest reliability (.90), while the measure of dangerous driving had the lowest (.49). Reliability for the remaining measures ranged from .70 through .76. Estimates for the three types of substance use are consistent with earlier analyses of these measures (O'Malley, Bachman, and Johnston 1983). It is understandable that the measure of dangerous driving would have the lowest reliability, since this behavior was assessed indirectly through reports of traffic tickets and accidents.

All estimates of paths from the general factors to the measured variables were highly significant (all $t > 6.2$), indicating that each behavior shared substantial variance with the others. Thus, these five deviant behaviors are potentially subject to some degree of shared explanation. Nevertheless, the importance of general deviance varied considerably across behaviors, and for some of the behaviors this proportion changed with time as well. Comparisons across behaviors are most straightforward in terms of the proportion of reliable variance (i.e., non-error variance) associated with the general and specific factors.

At all three times, criminal behavior was the form of deviance most closely associated with the general tendency, and dangerous driving was the behavior least associated. The proportion of reliable variance associated with the general factor ranged from 27 percent to 74 percent, but it rose above 50 percent only for criminal behavior.

There was little variation across time in the proportion of reliable variance in marijuana use and dangerous driving associated with the general factor. Both criminal behavior and heavy alcohol use became less associated with the general deviance over time. For criminal behavior, this trend accompanied a decline in the amount and variance of the behavior. Heavy

alcohol use remained equally prevalent but became more independent of other types of deviance as respondents reached legal drinking age. Over time, the use of illicit drugs other than marijuana became increasingly associated with other forms of deviance.

Longitudinal relationships. Estimates for paths among the latent variables in Model VIII appear in Figure 2. Figure 2 is comparable to Figure 1, expanded to five deviant behaviors and limited to the latent variables. There was considerable longitudinal stability for both general and specific factors. The lowest unstandardized coefficient reflecting stability was .53, and eight of the twelve were above .75.

Note that it is very unlikely that the stability of the specific factors would be due to memory effects such as reporting about the same incidents at more than one wave. The measures concerned behavior during the past year or less, and the interval between waves generally was two years.

Only one pair of longitudinal paths indicated influence between behaviors rather than stability of a behavior. These were the paths from earlier marijuana use to later use of other illicit drugs. This influence was of moderate size for the time-one to time-two interval (standardized beta of .27), but it was insignificant for the time-two to time-three interval (standardized beta of .09).

Table 3 expresses longitudinal influences in terms of the variance accounted for by measures at the preceding time.[9] Model VIII allows a separation of explained variance into general and specific components, using the original metric of each variable. This provides a more direct comparison of general and specific contributions than do the path coefficients reflecting stability. The specific variance explained is a function of the amount of specific variance at the previous wave and the stability of that variance (plus any influence from other specific behaviors). The general variance explained for a behavior is a function of both the variance that behavior shares with the general factor at the current wave and the stability of the general factor.

A large proportion of the reliable variance of all of the deviant behaviors can be explained by earlier measures of deviant behavior, with estimates ranging from 43 percent to 73 percent. Generally speaking, the proportion of reliable

---

[9] The variance explained by specific features of a behavior equals the amount of specific variance (see Table 2) minus the unexplained specific variance (psi for that latent variable). Wave one data explained 61 percent of the variance in general deviance at wave two (the square of the stability coefficient for the standardized latent variables), and wave two data explained 60 percent of the variance in general deviance at wave three.
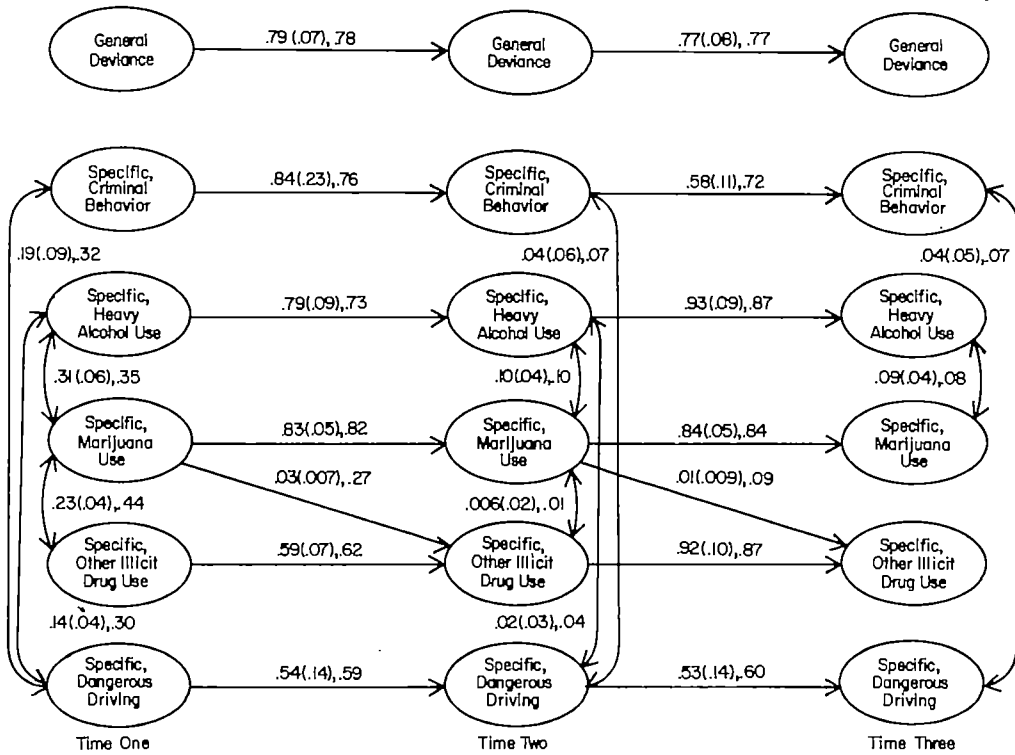
Fig. 2. Path Estimates for the Structural Model (Model VIII): Unstandardized Path Estimates Followed by their Standard Errors (in parentheses) and Standardized Path Estimates.

variance explained by general versus specific factors is as would be expected from the breakdown of the total variance into those two categories. Thus, most of the explained variance for criminal behavior is due to earlier general deviance, while most of the explained variance for marijuana use is due to earlier specific variance.

There were also some interesting variations in

the stability of the specific factors. The highest proportion of reliable variance explained was for heavy drinking at time three (71 percent) and use of illicit drugs other than marijuana at time three (73 percent). In both cases the specific variance component was extremely stable from time two to time three (unstandardized betas of .93 and .92). The next highest levels of stability were for marijuana use, the other form of

Table 3. Variance Explained by Longitudinal Influences

	Variance Explained by Preceding Wave			Percent of Reliable Variance Explained		
	General	Specific	Total	General	Specific	Total
Criminal behavior						
Time 2	2.81	2.23	5.04	33.5	26.5	60.0
Time 3	1.48	1.30	2.78	29.7	26.1	55.9
Heavy alcohol use						
Time 2	.30	.43	.73	23.4	33.1	56.6
Time 3	.24	.68	.92	18.6	52.3	70.9
Marijuana use						
Time 2	1.54	3.08	4.62	21.6	43.3	64.9
Time 3	1.51	3.24	4.75	21.3	45.7	66.9
Other illicit drug use						
Time 2	.027	.035	.062	26.3	34.4	60.7
Time 3	.034	.055	.089	27.7	44.9	72.6
Dangerous driving						
Time 2	.13	.16	.29	19.3	23.7	43.1
Time 3	.09	.13	.22	17.0	25.7	42.8

substance use, and its stability did not change over time (unstandardized betas of .83 and .84).

DISCUSSION

Our findings concerning general and specific features of deviant behaviors lead to a mixed conclusion about the generality of deviance. Hirschi (1984) and Jessor and his colleagues (Donovan and Jessor 1985; Jessor and Jessor 1977) were correct that a general tendency toward deviance could explain the positive correlations between different deviant behaviors. Indeed, a single latent variable can account for virtually all of their cross-sectional and longitudinal relationships. Nevertheless, a latent variable of general deviance falls far short of explaining all of the reliable and stable variance of the separate behaviors. Therefore, a theory that addresses only the general construct can never fully account for the separate behaviors, though it might account for much of each of them. Each behavior is, in part, a manifestation of a more general tendency and, in part, a unique phenomenon.

It may still be possible to explain all of the behaviors within a unified framework, such as each behavior resulting from peer-group norms for that particular behavior or all forms of deviance being influenced (but not totally determined) by a general alienation. Even so, factors important to one behavior could be entirely irrelevant to others, as illustrated by Johnston's (1973) finding that support for the counterculture during the Vietnam era was strongly correlated with certain types of illicit drug use, but not correlated with delinquency.

Our analysis also provided a test for the influence of specific deviant behaviors on one another. Marijuana use during the high school senior year had significant impact on use of other illicit drugs one to two years later. During the subsequent two years, the influence of marijuana use on later use of other illicit drugs was negligible. This result suggests that any influences of one behavior on another are age-specific, perhaps depending on age-related role transitions.

While this instance of an effect of one behavior on another is of interest, it is more important for our understanding of deviance that influence of this type was so limited. Only one of the twenty possible paths between these five behaviors was statistically significant. Though this does not appear to be a chance relationship, it yielded a negligible improvement in the overall fit of the model. It is clear that influences between specific behaviors contribute very little to the general finding that people who engage in one form of deviance are likely to engage in others as well.

Involvement in one form of deviant behavior is predictive of later involvement in others, not because of mutual influences, but because each partially reflects a general tendency toward deviance. For instance, frequent drunkenness in the senior year of high school would indicate a willingness to violate conventional standards of behavior. Since the general tendency toward deviance is relatively stable over time, this willingness is likely to become manifest in other forms of deviant behavior in the following years, as well as in the persistence of heavy drinking.

Our analysis of the general and specific components of each of the five behaviors yielded interesting insights into their shifting associations over time. Criminal behavior proved to be the type of deviance most closely linked to the general tendency, though the strength of the tie declined over time, as did the rate and variance of this behavior. Alcohol use among high school students was more strongly associated with a general willingness to flout conventional mores than was alcohol use among adults in their early twenties. The opposite was true for use of illicit drugs other than marijuana, which became increasingly tied to general involvement in deviance. Though their relations to general deviance diverge, each of these more serious forms of substance abuse had unique aspects that became extremely stable during respondents' early twenties.

Since deviance is defined by conventional standards for behavior, we would expect the overlap between general deviance and any particular behavior to fluctuate with variations in those standards. This is illustrated by our finding that heavy alcohol use is less related to general deviance once respondents reach the legal drinking age. Further research might apply our conceptual and analytic framework to additional tests of this proposition. For instance, we would expect cigarette smoking and sexual activity to be highly related to general deviance during early adolescence because these behaviors are considered inappropriate at this age. As these activities become more acceptable in late adolescence and early adulthood, they should be less connected to the general syndrome of deviance. In a similar vein, cross-cultural and temporal comparisons would provide a basis for testing whether normative standards influence the strength of the connection between a behavior and general deviance.

As with any piece of research, there are limitations to our study that should be addressed by future work. We offer a picture of the generality of deviance derived from samples of 18- to 22-year olds, all of whom had remained in high school until late in the senior year. Furthermore, while the five behaviors we have

considered are prominent for this age group, they hardly exhaust the meaning of the larger category of deviance.

Though our findings indicate considerable generality and stability of deviance, it has yet to be determined whether this pattern holds across the full range of deviant behaviors. As with much of the research in this area, we have concentrated on the problem behaviors of adolescence and early adulthood. It is not at all clear that deviance among adults, particularly middle-class adults, would fit the same pattern. Is the scientist who falsifies research results also more likely to cheat on taxes, be unfaithful to his or her spouse, and get into fights at bars? And do adolescent behaviors such as petty theft, illicit drug use, and dangerous driving predict a broad range of adult deviance? There is much to learn about the generality of deviance.

REFERENCES

Akers, Ronald L. 1977. *Deviant Behavior: A Social Learning Perspective*. Belmont, CA: Wadsworth.
_____. 1984. "Delinquent Behavior, Drugs, and Alcohol: What is the relationship?" *Today's Delinquent* 3:19–47.
Bachman, Jerald G. and Lloyd D. Johnston. 1978. *The Monitoring the Future Project: Design and Procedures*. (Occasional Paper #1.) Ann Arbor, MI: Institute for Social Research.
Bachman, Jerald G., Patrick M. O'Malley, and Jerome Johnston. 1978. *Adolescence to Adulthood: Change and Stability in the Lives of Young Men*. Ann Arbor, MI: Institute for Social Research.
Bentler, Peter M. and Douglas G. Bonett. 1980. "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures." *Psychological Bulletin* 88:588–606.
Billy, John O.G. and J. Richard Udry. 1985. "Patterns of Adolescent Friendship and Effects on Sexual Behavior." *Social Psychology Quarterly* 48:27–41.
Cloward, Richard A. and Lloyd E. Ohlin. 1960. *Delinquency and Opportunity: A Theory of Delinquent Gangs*. New York: Free Press.
Coleman, James S. 1961. *The Adolescent Society*. New York: Free Press.
Donovan, John E. and Richard Jessor. 1985. "Structure of Problem Behavior in Adolescence and Young Adulthood." *Journal of Consulting and Clinical Psychology* 53:890–904.
Elliott, Delbert S. and David Huizinga. 1984. *The Relationship Between Delinquent Behavior and ADM Problems*. Boulder, CO: Behavioral Research Institute.
Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton. 1985. *Explaining Delinquency and Drug Use*. Beverly Hills, CA: Sage.
Gibbs, Jack P. 1981. *Norms, Deviance, and Social Control: Conceptual Matters*. New York: Elsevier.
Gold, Martin. 1970. *Delinquent Behavior in an American City*. Belmont, CA: Brooks/Cole.
Gottfredson, Michael and Travis Hirschi. 1986. "The True Value of Lambda would Appear to be Zero: An Essay on Career Criminals, Criminal Careers, Selec-

tive Incapacitation, Cohort Studies, and Related Topics." *Criminology* 24:213–33.
Hays, Ron D., Keith F. Widaman, M. Robin DiMatteo, and Alan W. Stacy. 1986. "Structural-equation Models of Current Drug Use: Are Appropriate Models so Simple(x)?" *Journal of Personality and Social Psychology* 52:134–44.
Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley, CA: University of California Press.
_____. 1984. "A Brief Commentary on Akers' 'Delinquent Behavior, Drugs, and Alcohol: What is the Relationship?' " *Today's Delinquent* 3:49–52.
Hirschi, Travis and Michael Gottfredson. 1983. "Age and the Explanation of Crime." *American Journal of Sociology* 89:552–84.
_____. 1986. "The Distinction Between Crime and Criminality." In *Critique and Explanation: Essays in Honor of Gwynn Nettler*, edited by Timothy Hartnagel and Robert Silverman. New Brunswick, NJ: Transaction.
Jackson, Elton F., Charles R. Tittle, and Mary J. Burke. 1986. "Offense-Specific Models of the Differential Association Process." *Social Problems* 33:335–56.
Jessor, Richard and Shirley L. Jessor. 1977. *Problem Behavior and Psychosocial Development: A Longitudinal Study of Youth*. New York: Academic Press.
Johnston, Lloyd D. 1973. *Drugs and American Youth*. Ann Arbor, MI: Institute for Social Research.
Johnston, Lloyd D., Jerald G. Bachman, and Patrick M. O'Malley. 1986. *Monitoring the Future: Questionnaire Responses from the Nation's High School Seniors, 1985*. Ann Arbor, MI: Institute for Social Research.
Johnston, Lloyd D., Patrick M. O'Malley, and Jerald G. Bachman. 1986. *Drug Use Among American High School Students, College Students, and Other Young Adults: National Trends Through 1985*. Washington, DC: National Institute on Drug Abuse.
Johnston, Lloyd D., Patrick M. O'Malley, and L.K. Eveland. 1978. "Drugs and Delinquency: A Search for Causal Connections." Pp. 137–56 in *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues*, edited by Denise B. Kandel. Washington, DC: Hemisphere.
Joreskog, Karl G. and Dag Sorbom. 1978. *LISREL IV: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Chicago, IL: National Educational Resources.
Kandel, Denise B. 1978. "Homophily, Selection, and Socialization in Adolescent Friendships." *American Journal of Sociology* 84:427–36.
Kandel, Denise B., Ronald C. Kessler, and Rebecca Z. Margulies. 1978. "Antecedents of Adolescent Initiation into Stages of Drug Use: A Developmental Analysis." Pp. 73–99 in *Longitudinal Research on Drug Use: Empirical Findings and Methodological Issues*, edited by Denise B. Kandel. Washington, DC: Hemisphere.
Kaplan, Howard B. 1975. *Self-Attitudes and Deviant Behavior*. Pacific Palisades, CA: Goodyear.
Kessler, Ronald C. and David F. Greenberg. 1981. *Linear Panel Analysis: Models of Quantitative Change*. New York: Academic Press.
Krosnick, Jon and Charles M. Judd. 1982. "Transitions in Social Influence at Adolescence: Who Induces Cigarette Smoking?" *Developmental Psychology* 18:359–68.

Lemert, Edwin M. 1972. *Human Deviance, Social Problems and Social Control,* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

O'Malley, Patrick M., Jerald G. Bachman, and Lloyd D. Johnston. 1983. "Reliability and Consistency in Self-Reports of Drug Use." *International Journal of the Addictions* 18:805–24.

Osgood, D. Wayne, Lloyd D. Johnston, Patrick M. O.'Malley, and Jerald G. Bachman. 1985. *Monitoring the Future: Follow-Up Study, Form 2, High School Senior Classes of 1976–1980, Waves 1–3* [MRDF]. Ann Arbor, MI: Survey Research Center, The University of Michigan [producer].

Sutherland, Edwin H. and Cressey, Donald R. 1955. *Principles of Criminology,* 5th ed. Philadelphia: Lippincott.

THE SOCIAL ORGANIZATION OF SELF-HELP:
A STUDY OF DEFENSIVE WEAPON OWNERSHIP*

Douglas A. Smith
Craig D. Uchida
University of Maryland

Using Donald Black's theory of self-help, and drawing on the collective security model and the "fear and loathing" hypothesis, this paper examines factors influencing defensive weapon ownership. Interview data from 9,021 randomly selected households in 59 residential neighborhoods are used to assess the association between household and neighborhood characteristics and decisions to purchase weapons for self-protection. Findings indicate that defensive weapon ownership varies inversely with other forms of social control, such as the perceived effectiveness of police. Additionally, purchases of defensive weapons are associated with household demographics, past victimization, and perceived risk of crime. Collectively, these results indicate that self-help (defensive weapon ownership) is a structured form of social action.

INTRODUCTION

Self-help is a decentralized form of social control. This form of social control is evident, for example, when people take individual measures to protect themselves from the threat of crime. Citizens may purchase weapons for self-protection or install security devices in their homes. Such acts are instances of self-help (see Black 1980, 1983).

Research reported in this paper examines individual and contextual factors associated with decisions to purchase a gun or other weapon for self-protection. For example, does being a victim of crime increase the likelihood of self-protective weapon purchases? Does the crime rate of neighborhoods independently influence residents' decisions to purchase weapons for self-protection? Does the prevalence of self-help vary with other forms of social control?

During the past decade, a number of scholars have attempted to explain gun ownership with such factors as crime rates, fear of crime, and a host of other variables. Few firm conclusions about which factors explain weapon ownership have emerged from these studies. As a result, the widely held belief that Americans purchase weapons because of fear of crime remains inconclusive and controversial. For example, Wright, Rossi, and Daly (1983) have recently argued that no credible evidence exists to support popular claims that fear of crime increases weapon ownership.

A clearer understanding of self-protective weapon ownership has been inhibited by methodological and conceptual problems. First, many studies of weapon ownership do not distinguish among the various reasons why people own guns (e.g., protection, sport and recreation, or as collectors). This is especially critical given Wright's finding that "sport ownership exceeds ownership for reasons of self-protection among all gun owners by a factor of 3 to 1" (1984, p. 305). Even among handgun owners, Wright reports that less than half give self-protection as the most important reason why they own a gun.

Thus, studies which do not differentiate the motivating reason for weapon ownership are fitting models to data using a conceptually confounded dependent variable. This results in weak tests. Specifically, if fear of crime is related to self-defensive weapon purchases but not related to weapon purchases for sport or recreation, then models that simply use gun ownership as the dependent variable are inconsistent with the underlying theory. Unless the dependent variable is constructed to differentiate between ownership motivated by self-protection and other reasons for owning weapons, the interpretation of model parameters will be problematic. This problem is addressed in the research reported in this paper.

A second limitation of research in this area is the fact that most studies only consider a limited range of potential explanatory variables. For example, the majority of research looks either at individual or household level data to explain why people own guns, or focuses exclusively on data aggregated at the city, state, or national level. Yet, the emerging theoretical explanations of self-protective weapon ownership often cut across levels of aggregation. For example, the

crime rates and racial composition of neighbor-hoods are two factors that may influence individual decisions to purchase weapons for self-protection.

Moreover, most research in this area has been driven by what Wright and his associates (1983) refer to as the fear and loathing hypothesis: People buy weapons to protect themselves because of increasing fear of crime. Recently, McDowall and Loftin (1983) have advanced a collective security hypothesis: people take steps to protect themselves when their faith in the ability of agencies such as the police to protect them diminishes. Using aggregate data from Detroit, McDowall and Loftin find support for this perspective. Yet, this theory is ultimately a theory of individual choice and must be examined with individual level data.

Thus, one goal of the current research is to examine an integrated theoretical model of decisions to purchase guns or other weapons for self-protection. Our perspective draws on Black's (1980) theory of self-help, the collective secu-rity model, and various versions of the fear and loathing hypothesis. Our unifying conceptual model relies heavily on the concept of vulnera-bility, and our analysis uses data at both the household and neighborhood levels.

SELF-HELP AND THE COLLECTIVE SECURITY MODEL

Black argues that self-help "is a social practice that has been commonplace in many settings and which is present to some degree nearly every-where" (1980, p. 194). As a quantitative variable, self-help varies with other dimensions of social life. Using history as a guide, Black notes that the quantity of self-help is highest in simplest societies and declines with the growth of law and formal institutions of social control. In general, then, self-help is a form of nonlegal social control that varies inversely with the amount of governmental social control.

Instances of self-help are many and varied. Organized vigilantism is a form of self-help (see Brown 1976). Additionally, citizens may at-tempt to protect themselves from crime through a variety of self-precautions, including weapon purchases, and this too is a form of self-help (see Lavrakas 1985). Finally, the growth of the private security industry reflects the desires of individuals and organizations for better protec-tion than that afforded by the police. As Sherman points out: "few developments are more indicative of public concern about crime—and declining faith in the ability of public institutions to cope with it—than the burgeoning growth of private policing" (1982, p. 87).

The prevalence of self-help is inversely related to the quantity and quality of formal social con-trol (Black 1980). With the growth of the police, citizens have become more dependent on gov-ernment to define and maintain order (Silver 1967). Reiss and Bordua (1967) have argued that citizen dependence on police for protection is a central component of maintaining order in mod-ern society. Moreover, they note that, when peo-ple perceive the police to be ineffective, they may take measures to protect themselves. If cit-izens feel secure from the risk of crime, self-help will be less prevalent. Conversely, if citizens perceive governmental control to be weak or in-effective, they will take steps to protect them-selves. Thus, we hypothesize that factors that measure individual vulnerability to crime, whether perceived or real, will increase the likelihood of citizens buying guns or other weapons for self-protection. One major component of vulnerabil-ity is confidence in the police.

Perhaps the most sophisticated statement of the thesis that public confidence in the police is a major factor explaining the demand for legal handguns is offered by McDowall and Loftin (1983). Their collective security model is based on a premise that self-interest engenders self-protection and closely parallels Black's self-help model, in that both assume that a weakening in governmental control stimulates self-protective action. McDowall and Loftin used the handgun license series in Detroit from 1955 to 1977 as their indicator of collective security, and in-cluded the violent crime rate, the riot of 1967, and the number of uniformed police officers per 100,000 residents as independent variables of pri-mary theoretical interest. Other explanatory vari-ables, such as race, age, and income composi-tion of the population, were also included. Their results show that Detroit residents "purchased more handguns when violent crime and civil dis-order reduced confidence in collective security and fewer when these conditions abated" (1983, p. 1,156). Additionally, citizens purchased fewer weapons when police strength was rising and bought more guns when it declined.

Thus, in examining the propensity of individ-uals to purchase guns or other weapons for protection, we hypothesize that conditions of vulnerability will play an important role in these decisions. One aspect of vulnerability is a perception that police are unable to provide effective security against crime. As this percep-tion rises, so too will instances of self-help. Other aspects of vulnerability, such as fear of crime, past victimization experience, and the crime rates of areas where people live, are also potentially important determinants of self-help and are discussed in the next section.

PRIOR STUDIES OF WEAPON OWNERSHIP

Most empirical studies of gun or weapon ownership are descriptive and tend to focus on

self-defense and self-protection as the reasons for ownership. While these studies do not directly deal with the notion of self-help, they address some of its more important components.

At the core of debates about gun ownership is the claim that fear of crime, violent crime, and civil disorder underlie the growth in weapon possession. Wright et al. (1983) argue, however, that this fear and loathing hypothesis lacks empirical support. Citing difficulties with a number of studies (e.g., Newton and Zimring 1969), Wright and his associates conclude that "there is *no* credible study *anywhere* in the literature that shows, clearly and unmistakably, a fear and loathing effect in the weapons trend" (1983, p. 101).

A review of major studies generally supports this assessment of the research. Empirical studies based on the National Opinion Research Center's General Social Survey (DeFronzo 1979; Stinchcombe, Adams, Heimer, Schepple, Smith, and Taylor 1978; Williams and McGrath 1978; and Wright and Marston 1975) have found that the criminally victimized are not any more likely than the nonvictimized to possess a weapon. Additionally, there is evidence that fear of crime is also not directly or strongly linked to gun ownership. Wright and Marston (1975) found that people who expressed some fears about going out alone at night were somewhat less likely to own a gun than those who did not. Weapon ownership was also disproportionately low among people who had been burglarized or threatened with force in the previous year. Williams and McGrath (1976), using 1973 NORC data, found that those who had been previously victimized were not any more likely to own guns than those who had not been victims of crime. They also reported a negative relationship between fear in the neighborhood and gun ownership. DeFronzo (1979), using data from the 1973, 1974, 1976, and 1977 General Social Surveys, found results similar to his predecessors: victimization and the fear of crime had no significant effect on handgun ownership.

While these studies reveal little association between fear of crime or victimization and gun ownership, McDowall and Loftin (1983) point out that each study is problematic, either because of flaws in the analyses or because the underlying models were incompletely specified (McDowall and Loftin 1983, p. 1,149). Moreover, results from other research indicate that it may be premature to discard the fear and loathing hypothesis. Bordua and Lizotte (1979), using the incidence of Firearms Owners Identification Cards in Illinois counties found that the country crime rate was a significant factor influencing female gun ownership. In subsequent work, Lizotte, Bordua, and White (1981) found that victims of violent crime were significantly more likely to purchase guns for self-protection than were nonvictims. Additionally, those more fear-

ful of crime were significantly more likely to own guns for self-protection than those who were less fearful. At the same time, they found that self-defensive weapon ownership was not directly influenced by proximity to blacks or by county crime rates. More recently, Hill, Howell, and Driver (1985) reported that both fear of crime and prior victimization significantly influence self-protective weapon ownership among males. Taken together, these studies indicate that the fear and loathing hypothesis may be more inconclusive than incorrect.

THE CURRENT RESEARCH

Based on previous research this paper examines a model of weapon ownership for self-protection that interprets the relationship between household and neighborhood characteristics and weapon ownership within a framework of self-help. Using data on individual households and neighborhood attributes, we hope to identify those aspects of neighborhoods (such as crime rates, racial composition, and economic status) and households (such as sex, race, and income composition, prior victimization experience, and perceptions of police) that are associated with decisions to purchase guns or other weapons for self-protection. By focusing on weapon ownership, self-help becomes a quantitative variable that may be greater under some conditions and weaker under others. In the next section we describe the data used in this research and the operationalization of variables.

DATA AND VARIABLES

The data used in this research come from 9,021 interviews with random samples of residents in 59 residential neighborhoods located in three Standard Metropolitan Statistical Areas (Rochester, NY; Tampa-St. Petersburg, FL; and St. Louis, MO).[1] These data were collected in 1977 as part of the Police Services Study.[2]

---

[1] Neighborhoods are defined on the basis of police-beat boundaries, census block groups, and enumeration districts. They are primarily residential areas, ranging in size from .2 to 3.8 square miles with an average size of 1.5 square miles. Neighborhood areas were selected to represent variation on racial and income lines. Areas range from all white to all black and median household income varies from $4,208 to $22,315. The original sample consisted of 12,019 interviews. Of these, 221 were excluded by a decision to drop the data from one neighborhood that was an extreme outlier on certain neighborhood variables that will be used in this analysis. The additional lost cases resulted from missing data, which was uniformly distributed across the remaining 59 neighborhoods.

[2] The Police Services study was conducted in the summer of 1977 under the auspices of the Workshop in Political Theory and Policy Analysis at Indiana Univer-

As noted earlier, one of the major problems in the literature on weapon ownership is the conceptual ambiguity of the dependent variable. While most studies acknowledge that handguns have multiple uses, prior analyses have often failed to distinguish among the various reasons why people own guns. Recreation (sport), investments, and protection against both animals and people are a few of the major reasons why people own weapons (see Wright et al. 1983). Moreover, Wright (1984) and Kleck (1979) have shown that 60 percent of handguns are owned for reasons other than self-defense. Thus, studies that construct the dependent variable to contrast all owners with nonowners may seriously distort the effects of independent variables, such as fear of crime, on weapons ownership for reasons of protection. For example, Wright and Marston (1975), Williams and McGrath (1976), and DeFronzo (1979) used all gun owners from the NORC survey to test the fear and loathing hypothesis. But theory suggests that fear of crime increases *self-protective* weapon ownership but not ownership of weapons for other reasons. The failure to incorporate this distinction in prior empirical studies has lead to inconclusive results.

Research reported in this paper uses a dependent variable that clearly identifies households that have purchased weapons for self-protection. As part of the interview, respondents were read the following statement: "People nowadays are taking precautions to protect their home and families from crime. For the next few questions please tell me whether you or any member of your household have done the following things to increase your safety from crime." One of the subsequent questions read: "Have you purchased a gun or other weapon for your protection?" Our analysis will focus on responses to this question, which are coded one for yes and zero for no. In this sample, 14.3 percent of the households reported purchasing a gun or other weapon for protection.

Respondents acted as informants for a household. Thus, our primary unit of analysis is the household and not individuals. Several variables were constructed to measure household attributes. These include racial composition of the household and the type of household: whether it is occupied by a family, a childless couple, a single parent, or a household consisting of only males or only females.[3] This coding of house-

hold type allows the best possible categorization given these data. Based on extant research, we expect to find a higher prevalence of self-protective weapons in households where a male is present. Additionally, data are available on household income, which is strongly related to weapons ownership in several prior studies.

In addition to household demographic data, several questions about crime in the area and the respondents' perceived chances of being victimized were included in these surveys. For example, respondents were asked whether they felt that crime was increasing in their neighborhood (defined as the two or three blocks around their house).[4] Additionally, respondents were asked to estimate the chances that their household would be burglarized or that they would be robbed in the neighborhood in the next year. If the respondents reported that they perceived either of these events as "very likely," they were coded as having a high perceived crime risk. Finally, respondents were asked if they or any member of their household had been victimized in the last 12 months. If they were the victim of a personal or property crime (excluding vandalism), they were coded as being a prior victim.

Information to address one aspect of the collective security thesis is also available in these data. Respondents were asked to rate the quality of police services in their neighborhoods. Responses ranged along a five-point Likert scale from "very poor" to "outstanding." The self-help model predicts that prevalence of defensive weapon ownership will decline as one's rating of police effectiveness increases.

Finally, since these data were collected in 59 residential neighborhoods, we have aggregated several household variables up to the neighbor-

[3] These categories are created using information on the age and sex of each resident in a household For example, a family is defined as any household with at least one male and one female adult and one child. Couples' households are defined as homes occupied by two adults of different gender groups. All single-parent households contain a female parent (a few male-parent cases were excluded from this analysis). Male and female households are those occupied exclusively by males or females.

[4] Since only one person was interviewed in each sampled household, the degree to which their reported perceptions represent the views of other household members may be problematic, and some caution is warranted in interpreting the results pertaining to these variables. However, this caveat applies to variables where it is reasonable to expect that household members share more similar than dissimilar views. For example, adult members of households are probably more likely to agree than disagree about whether crime in their immediate area has been increasing in the past year. Similarly, it seems reasonable to expect that household members will hold more similar than different views about the quality of police services in their area. In the absence of interview data from multiple members of the same household, certain results must be interpreted cautiously.

hood level. These include the victimization rate, racial composition, percent of single parent households, and percent of households with annual incomes of less than $5,000. In the following analysis we will examine whether decisions to purchase guns or other weapons for self-protection are independently influenced by attributes of the areas where people live.

FINDINGS

We begin our analysis by reporting the percentage of households that have purchased a weapon for self-protection by categories of independent variables at the household level. These data are reported in Table 1 and are generally consistent with our expectations. For example, households whose members have been victimized in the past year are more likely to own weapons for protection than households in which no members have been the victims of crime. Addition-

Table 1. Self-Help by Household Characteristics (N = 9,021)

Variable	N	Percent with Weapon for Self-Protection
All Cases	9,021	14.3
Race		
White	6,438	13.7
Nonwhite	2,583	15.7
Household Type		
Family	4,508	16.3
Couples	2,230	13.7
Single-Parent	552	9.2
Male-Only	572	19.2
Female-Only	1,159	7.2
Household Income		
Below $5,000	1,753	10.2
$5,001–$10,000	2,105	13.6
$10,001–$15,000	1,919	15.1
$15,001–$20,000	1,488	16.3
$20,001–$25,000	848	15.0
$25,001–$30,000	430	17.7
Above $30,000	478	17.8
Rating of Police		
Very Poor	316	24.4
Inadequate	571	21.9
Adequate	2,784	14.8
Good	4,199	12.6
Outstanding	1,151	12.3
Victim Status		
No	6,731	12.3
Yes	2,290	20.0
High Perceived Crime Risk		
No	7,619	13.1
Yes	1,402	20.4
Crime Increasing		
No	6,968	12.9
Yes	2,053	18.8
Home Ownership		
Renter	1,961	11.5
Owner	7,060	15.0

ally, households in which the respondents perceive a greater risk of future victimization or saw the crime rate of their neighborhood increasing in the past year were more likely to own defensive weapons than households where respondents did not share these views. Thus, three crime-related measures of vulnerability are each associated with increased self-help.

A second dimension of vulnerability is faith in the police to protect the lives and property of individuals. These data indicate that, in households where the respondents evaluated police services as very poor, 24.4 percent owned guns or other weapons for self-protection, compared to 12.3 and 12.6 percent of households that rated police services as outstanding or good respectively. Indeed the prevalence of self-help declines monotonically as ratings of police services improve.

Some past research has noted that those with more financial resources are more likely to own guns. The current data also indicate that, as household income increases, so does the likelihood of owning a weapon for self-protection. Also consistent with past research is the finding that self-protective weapon ownership varies as a function of gender. Households occupied by only males are the most likely to own a weapon for self-protection (19.2 percent) while those composed of only females are least likely to own such weapons (7.2 percent). Finally, self-help does not vary appreciably with race.

This admittedly brief overview of the bivariate association between self-protective weapon ownership and several of our independent variables is intentional. Our primary focus in this paper involves assessing the *relative* contribution of these variables and neighborhood characteristics to decisions to purchase weapons for self-protection. A multivariate method is necessary to determine which factors influence self-help independently of other variables in the model. Moreover, we are interested in two other issues. First, are the effects of independent variables on defensive gun ownership conditional on other measured variables? Second, we want to examine whether neighborhood characteristics independently influence decisions to purchase weapons for protection. To make these assessments we employ a logistic regression model.

The primary results of our analysis are reported in four equations presented in Table 2. The first equation corresponds to a model in which only household characteristics are allowed to influence weapon ownership for self-protection. The second equation reports final results from a model in which the effects of household characteristics are allowed to affect weapon ownership conditional on other household characteristics. The third equation reported

Table 2. Logistic Regression Models of Self-Help with Household and Neighborhood Characteristics ($N = 9,021$)

Independent Variables	2.1		2.2		2.3		2.4	
	$B*$	z-value	$B*$	z-value	$B*$	z-value	$B*$	z-value
Couple	−.029	−.37	−.031	−.41	−.034	−.43	−.038	−.49
Single-Parent Household	−.648	−4.09	−.618	−3.91	−.634	−4.01	−.639	−4.04
Male-Only	.372	3.10	.378	3.15	.355	2.96	.383	3.18
Female-Only	−.642	−4.87	−.634	−4.92	−.643	−4.99	−.642	−4.99
Nonwhite Household	.181	2.53	.185	2.59	.061	.74	.096	1.15
Household Income	.069	3.44	.039	1.77	.050	2.24	.058	2.57
Rating of Police	−.148	4.35	−.146	4.30	−.134	−3.90	−.122	−3.53
Home Ownership	.288	3.48	.281	3.34	.313	3.67	.252	2.92
Crime Victim	.407	5.89	.402	5.80	.386	5.57	.388	5.57
Crime Risk	.291	3.52	−.238	1.41	−.265	−1.57	−.240	−1.42
Crime Trend	.255	3.52	.252	3.48	.247	3.40	.281	3.84
Crime Risk × Household Income			.172	3.68	.175	3.73	.170	3.63
Percent Single-Parent Households					.044	3.14	.041	2.87
Percent Population Aged 12–20					−.004	.47	.005	.51
Site Effect—Tampa							.288	3.92
Site Effect—Rochester							−.023	.24
Intercept		−1.941		−1.840		−2.039		−2.349
Likelihood Ratio		254.02		267.54		278.27		298.00

* Maximum Likelihood Logit Coefficient.

in Table 2 presents final results from a model that introduces neighborhood characteristics as independent variables. Finally, the fourth equation includes two dummy variables that capture the covariation between the three SMSAs and self-defensive weapon ownership that is not explained by other variables in the model.

Results from equation 2.1, reported in Table 2, are generally consistent with the bivariate associations reported above. For example, all three crime-related variables independently influence the probability of owning a weapon for self-protection. We should note that in a sample of this size there is a bias toward finding statistically significant effects, since the standard errors of the logit coefficient are attenuated by sample size. However, point estimates of logit coefficients are not influenced by sample size and can be manipulated to reflect the influence of the independent variables on changes in the marginal probability of owning a weapon for self-protection. Consider prior victimization. The logit coefficient is .407. If we ask how much does being a victim of crime increase the probability owning a weapon for self-protection, we need to establish a baseline probability of ownership. Conceptually, this baseline probability of ownership is the probability of ownership for persons who have not been victimized. If the baseline probability of ownership was .143 (the average probability of ownership in the sample), then being a victim of crime would increase the estimated probability

of ownership to .201 (a 40.5 percent increase).[5] We would emphasize that these estimates of the effect of victimization on weapon ownership are averages for all victims and are not influenced by sample size.

Considering household composition, the estimates reported in equation 2.1 are relative to the reference category (families). These results indicate that female-only and single-parent households are less likely to own weapons for self-protection than families. The estimated probabilities of ownership for female-only and

[5] These figures were estimated in the following manner. The predicted value in a logit model ($\hat{Y}$) can be converted into an estimated probability (p) by $p = 1 / (1 + e - \hat{Y})$, where $e = 2.718$. If $\hat{Y} = -1.79$, then $p = .143$, which is the sample mean. Thus, a household that had a predicted value of $\hat{Y}$ of 1.79 would have an estimated probability of owning a weapon for defensive purposes of .143. Using this as the baseline probability of ownership, we can estimate the probability of ownership for a household that was similar to the baseline household in all respects except one (i.e., a household member has been a victim of crime). The effect of prior victimization from equation 2.1 is .407. Thus, the $\hat{Y}$ for this household would be $-1.79 + .407 = -1.383$. Using the above expression, this would translate into an estimated probability of ownership of .201. Some authors prefer to estimate changes in the probability of a dependent variable from an arbitrary baseline probability of .5. We believe that this overstates the influence of independent variables on the probability of a dependent variable, since the logistic function is changing most rapidly at this point. By using the sample mean of weapon ownership as the baseline probability, our estimates of the influence of independent variables on changes in the probability of weapon ownership are conservative.

single-parent households are .081 and .080, respectively, using a baseline probability of ownership of .143. Alternatively, households consisting of only males are more likely to engage in self-help relative to households occupied by families (estimated probability of ownership = .196). Collectively, these results indicate that self-protective weapon ownership varies with the presence of males in the household. Moreover, these gender differences are independent of other variables included in the model and thus cannot be explained by differences between males and females with respect to these variables. These results suggest that self-help may contain a subcultural or socialization component that is not directly measured in the current data.

Other results from equation 2.1 also confirm the bivariate associations. As household income rises, so does the likelihood of self-help. Additionally, as ratings of police service become more favorable, the probability of self-protective weapon ownership declines. Finally, home owners are more likely to engage in self-help relative to those who rent.

Our next focus concerns whether the associations revealed by equation 2.1 are additive or conditional. To address this issue, we estimated exploratory models testing several theoretically motivated conditional relationships. One significant interaction involving household income and perceived crime risk was found. The results of the model with this interaction term are reported as equation 2.2 in Table 2. Once this interaction is included in the model, no other interaction term emerged that significantly influenced the probability of self-protective weapon ownership.

Results from equation 2.2 indicate that the association between income and self-help varies by how household members assess their risk of future victimization. This pattern is best indicated by the data presented in Table 3, which reports actual probabilities of weapon ownership by income levels and perceived crime risk. For households in which respondents report a low risk of future victimization, increasing income is only weakly associated with defensive weapon ownership. However, in households where respondents believe their chances of being

victimized are high, defensive weapon ownership increases markedly with rising income. For example, over 4 out of 10 households with annual incomes greater than $30,000 and that perceive their crime risk to be high have purchased a gun or other weapon as a defense against crime, compared to only 12 percent of households in the lowest income category. Moreover, the last two rows of Table 3 present actual victimization probabilities by household income and indicate that variation in self-help by income responds more to perceived than actual victimization risk.

Our next question concerns whether characteristics of neighborhoods in which people live affect their decisions to purchase weapons for protection, once we control for the effects of household level variables. In addition to the victimization rate of the neighborhood, four characteristics of neighborhoods, drawn from the literature on the social ecology of crime, were selected for study. These include 1) the percent of households with annual incomes under $5,000; 2) the percent of nonwhite households; 3) a measure of residential mobility; and 4) the percent of single-parent households.

To address the question of whether neighborhood characteristics influence decisions to purchase weapons for protection, we initially estimated a logit model in which weapon ownership was regressed on the five neighborhood characteristics. In this model the victimization rate of the neighborhood, the percent of single-parent households in the neighborhood, and the percent of nonwhite households had significant positive effects on self-help. Next, we added the household variables to this equation to determine whether these three neighborhood effects were mediated by household characteristics. In this analysis, only the percent of single-parent households remained significant. The model was reestimated including only the percent of single-parent households (and an age composition variable as a control) and household characteristics. Results from this model are reported as equation 2.3 in Table 2. Further analysis revealed that the crime rate and racial composition of the neighborhood were no longer significant when three crime-related household measures were included in the model.

Table 3. Actual Probabilities of Defensive Weapon Ownership by Household Income and Perceived Crime Risk and Victimization Probabilities by Household Income

	Annual Income in Thousands of Dollars						
	Below 5	5–10	10–15	15–20	20–25	25–30	30+
Perceived Crime Risk							
Low	.10	.13	.14	.14	.14	.15	.16
High	.12	.17	.22	.29	.22	.37	.43
Personal Victimization	.07	.06	.06	.04	.05	.04	.04
Property Victimization	.17	.20	.24	.24	.24	.25	.24

Specifically, the influence of neighborhood crime rates and racial composition on household decisions to purchase weapons is explained by the victimization experiences of individual households and the perceived risk of future victimization.

While individual experiences and crime-related perceptions mediate the relationship between crime rates and racial composition of neighborhoods and decisions to purchase weapons for self-protection, they do not explain the influence of the percent of single-parent households on defensive weapon purchases. Moreover, results reported in equation 2.3 indicate that it is not the single-parent households that are buying weapons. Instead, other types of households are more likely to purchase weapons for protection as the percent of single-parent households in the neighborhood rises. In the social ecology of crime literature, the single-parent household variable is often interpreted as an indicator of community disorder. To the extent this conceptualization is valid, these results indicate that disorder has a general effect on defensive weapon ownership that is independent of the household characteristics in this analysis, including one's own perceived risk of crime.

Finally, we estimated a model that included two dummy variables to control for site effects. These results are presented in equation 2.4 in Table 2 and indicate that households in the Tampa SMSA have a higher average probability of owning a weapon for self-protection than households in Rochester or St. Louis. This difference in the marginal probability of ownership by SMSA cannot be explained by other variables included in this analysis. Moreover, while these site effects add explanatory power to the model, they do not alter any of the relationships previously discussed.

CONCLUSIONS

Using a dependent variable that differentiates between ownership motivated by self-protection and other reasons for owning weapons, this research has examined the influence of household and neighborhood characteristics on self-defensive weapon ownership. Results of this study indicate that, when citizens feel vulnerable to crime or when they perceive police services to be ineffective, the likelihood of self-help (purchasing weapons or protection) increases. These findings diverge from some previous research, which has found that crime victims are not more likely to own weapons. We believe this difference is attributable to differential operationalizations of the dependent variable across studies. Results reported in this paper only concern defensive weapon purchases and

not weapon purchases more generally. Thus, the current findings are consistent with hypotheses derived from the collective security model and the fear and loathing thesis with respect to defensive weapon purchases.

Collectively, these findings suggest that self-help is a structured form of social action. Self-help, which manifests itself in weapon purchases, is associated with households occupied by males, perhaps reflecting early gender-specific socialization or prior military experience, which results in males viewing weapons as a more viable means of defense than females. Households that have been victims of crime, and those whose members perceive a greater risk of future crime, are also more likely to purchase weapons for self-protection. Thus, conditions of vulnerability stimulate demand for defensive weapons. The demand for defensive weapons also rises as confidence in the police deteriorates, suggesting that self-help varies inversely with the perceived effectiveness of more formal institutions of social control. Moreover, self-defensive weapon ownership is more prevalent in communities characterized by lower levels of informal social control of youth.

Income also plays a major role in weapon purchases for self-defense, though the influence of income is conditional on levels of perceived vulnerability. Households at the low end of the income distribution are equally likely to make defensive weapon purchases whether their members perceive their future crime risk to be high or low. However, at the higher end of the income distribution, households whose members perceive their future crime risk to be high are three times as likely to purchase weapons for self-protection than households whose members perceive their chances of future victimization to be low. Thus, it is the joint occurrence of high perceived crime risk and household income that influences self-protective weapon purchases.

This finding has interesting ramifications in terms of current gun control legislation efforts. A number of state statutes and proposed legislative bills seek to control weapon ownership through licensing, permits, tariffs, and additional taxation of weapons. These methods increase the cost of guns and, therefore, reduce their availability. However, this type of legislation discriminates against lower income households who may desire weapons for defensive purposes but cannot afford them. Increasing the cost of defensive weapons simply expands the existing inequalities evident in this research.

Finally, the influence of certain community characteristics on self-help is mediated by residents' perceptions of their own vulnerability. Such neighborhood attributes as crime rates, racial composition, and levels of residential mobility have no independent influence on

citizens' purchases of weapons for self-protection. This is not to suggest that levels of self-help will not vary with these characteristics of communities; they will. Rather, it suggests that these neighborhood characteristics influence defensive weapon purchases indirectly through individual citizens' perceptions of their own vulnerability to crime.

In summary, self-defensive weapon ownership is a form of self-help that varies with perceived vulnerability to crime, the perceived effectiveness of formal social control institutions such as the police, and the extent of informal social control of youth in a community. This research suggests that self-help is an alternative form of social control that varies systematically with other dimensions of social life.

REFERENCES

Black, Donald. 1980. *The Manners and Customs of the Police*. New York: Academic Press.
———. 1983. "Crime as Social Control." *American Sociological Review* 48:34–45.
Bordua, David J. and Allan J. Lizotte. 1979. "Patterns of Firearms Ownership: A Cultural and Situational Analysis of Illinois Counties." *Law and Policy Quarterly* 1:147–75.
Brown, Richard M. 1976. "The American Vigilante Tradition." Pp. 153–86 in *Violence in America: Historical and Comparative Perspectives*, edited by Hugh D. Graham and Ted R. Gurr. Beverly Hills, CA: Sage.
DeFronzo, James. 1979. "Fear of Crime and Handgun Ownership." *Criminology* 17:331–39.
Hill, G., F. Howell, and E. Driver. 1985. "Gender, Fear, and Protective Handgun Ownership." *Criminology* 23:541–52.
Kleck, Gary. 1979. "Capital Punishment, Gun Ownership, and Homicide." *American Journal of Sociology* 84:882–910.
Lavrakas, Paul. 1985. "Citizen Self-Help and Neighborhood Crime Prevention Policy." Pp. 87–115 in *American Violence and Public Policy: An Update of the National Commission on the Causes and Prevention of Violence*, edited by Lynn A. Curtis. New Haven: Yale University Press.

Lizotte, Alan J., David J. Bordua, and Carolyn S. White. 1981. "Firearms Ownership for Sport and Protection: Two Not So Divergent Models." *American Sociological Review* 46:499–503.
McDowall, David and Colin Loftin. 1983. "Collective Security and the Demand for Legal Handguns." *American Journal of Sociology* 88:1146–61.
National Opinion Research Center. General Social Survey Cumulative File 1972–1986. Study #ICPSR8609.
Newton, George D. and Franklin E. Zimring. 1969. *Firearms and Violence in American Life: A Staff Report to the National Commission on the Causes and Prevention of Violence*. Washington, DC: Government Printing Office.
Reiss, Albert J., Jr. and David J. Bordua. 1967. "Organization and Environment: A Perspective on the Police." Pp. 28–40 in *The Police: Six Sociological Essays*, edited by David J. Bordua. New York: Wiley.
Sherman, Lawrence. 1982. "Watching and Crime Prevention: New Directions for Police." *Journal of Contemporary Studies* 80:87–96.
Silver, Allan. 1967. "The Demand for Order in Civil Society: A Review of Some Themes in the History of Urban Crime, Police, and Riot." Pp. 1–24 in *The Police: Six Sociological Essays*, edited by David J. Bordua. New York: Wiley.
Stinchcombe, Arthur L., Rebecca Adams, Carol Heimer, Kim Scheppe, Tom W. Smith, and D. Garth Taylor. 1980. *Crime and Punishment: Changing Attitudes in America*. San Francisco: Jossey-Bass.
Williams, J. Sherwood and John H. McGrath, III. 1976. "Why People Own Guns." *Journal of Communication* 26:22–30.
———. 1978. "A Social Profile of Urban Gun Owners." Pp. 51–61 in *Violent Crime: Historical and Contemporary Issues*, edited by James A. Inciardi and Anne E. Pottieger. Beverly Hills, CA: Sage.
Wright, James D. and Linda L. Marston. 1975. "The Ownership of the Means of Destruction: Weapons in the United States." *Social Problems* 23:93–107.
Wright, James D., Peter H. Rossi, and Kathleen Daly. 1983. *Under the Gun: Weapons, Crime, and Violence in America*. New York: Aldine.
Wright, James D. 1984. "The Ownership of Firearms for Reasons of Self-Defense." Pp. 301–28 in *Firearms and Violence: Issues of Public Policy*, edited by Don B. Kates, Jr. San Francisco: Pacific Institute for Public Policy Research.

CLASS IDENTIFICATION OF MEN AND WOMEN IN THE 1970s AND 1980s*

NANCY J. DAVIS
DePauw University

ROBERT V. ROBINSON
Indiana University

Employing NORC data for 1974–1978 and 1980–1985, we evaluate three models of class identification for married men and women: (1) an "independence" model, in which one's own characteristics outweigh those of one's spouse; (2) a "sharing" model, in which equal weight is attached to one's own and one's spouse's characteristics; and (3) a "borrowing" model, in which one's spouse's characteristics are more important than one's own. In both decades, married men exhibit an independence model, in which only their own characteristics are salient, regardless of whether their wife works outside the home or not. From the 1970s to the 1980s, women shifted from a borrowing model to a sharing model. In the 1970s, employed or not, women attached more weight to their husband's than to their own characteristics. In the 1980s, employed women moved toward a sharing model, in which they take into account both their own and their husband's characteristics. An analysis of single men and women revealed movement away from a sharing model, in which both their own and their parents' characteristics are important, toward an independence model, in which only their own characteristics are salient. These changes are examined in light of economic and demographic shifts, political activism by women, and a possible trend toward greater individualism.

INTRODUCTION

The assumption that a family's class position derives from the husband's position in the labor market has dominated both sociological theories of class and popular conceptions of it. In the classical traditions that defined class in terms of the relations of production (Marx [1893] 1909) or these relations in combination with market position (Weber [1922] 1969), married women, who were not employed in large numbers until recent times, would seem to have been only indirectly connected to the class structure through their husbands. While Marx and Weber did not address the class position of nonemployed wives, many contemporary researchers have concluded that the position of wives in the class structure is derived from that of their husbands. Parkin (1971, pp. 14–15) and Giddens (1973, p. 288), while not endorsing this condition, argued that it is the social reality. Most recently, Goldthorpe (1983) justified using the husband's characteristics to assess a family's class position because husbands usually have longer work histories than wives and because husbands and wives have similar types of employment. But some feminist scholars have

argued that the individual, and not the family, should be the unit of stratification (Acker 1973), or that both spouses' characteristics should be considered in determining a family's class position (Britten and Heath 1983).

Studies of the subjective class identification of married men and women during the 1970s found that the husband as the sole determinant of the family's class position had a strong basis in popular thought. By regressing a married person's class identification (e.g., upper class, middle class, working class, lower class) on his or her own characteristics (e.g., education, occupational prestige, income) and spouse's characteristics, researchers determined which characteristics husbands and wives used to decide their class position. These studies found that husbands took their own characteristics into account much more than their wives' characteristics in placing themselves in the class structure (Felson and Knoke 1974; Rossi et al. 1974; Ritter and Hargens 1975; Hiller and Philliber 1978; Philliber and Hiller 1978, 1979; Van Velsor and Beeghley 1979; DeFronzo 1981a, 1981b; Jackman and Jackman 1983). About the model held by wives, there was less agreement. Ritter and Hargens (1975), Hiller and Philliber (1978), and Van Velsor and Beeghley (1979) found that employed wives considered both their husbands' and their own characteristics in determining their class position. Felson and Knoke (1974) and Rossi and his colleagues (1974) found that wives tended to take into account only their husbands' characteristics.

The differences among these studies appear to reflect different model specifications. The most sophisticated analysis of this topic—Jackman and Jackman's (1983) study of 1975 data—found that both employed and nonemployed wives tended to take into account their husbands' characteristics more than their own.

For both scholars and the general populace, ideas about men's and women's class position are rooted in the objective social structure. Marx ([1845–46] 1978, p. 155), Mannheim (1936), and others have posited a connection between objective social structures and consciousness. Certainly, throughout the nineteenth century and well into the twentieth century, married women were largely confined to the relations of *reproduction* of labor and rarely participated directly in the relations of *production*. In 1890, fewer than 5 percent of married women in the U.S. were employed and, as recently as 1940, only 16 percent were in the labor market (U.S. Bureau of the Census 1975, p. 133). Since nearly all conceptions of class take into account the individual's position in the workplace as a primary consideration (e.g., Marx [1893] 1909, pp. 1031–2; Weber [1922] 1969, pp. 24–26; Dahrendorf 1959; Giddens 1973; Wright 1976), nonemployed married women could be defined as outside the class structure or, more often, as linked to it indirectly through their husband's employment.

Long-term social and economic changes, however, lead us to question the suitability of this conceptualization for contemporary U.S. society. The percentage of employed married women has risen from less than a quarter in 1950 to over half in 1985 (U.S. Bureau of the Census 1985, p. 398). The movement of many women into the workplace has been accompanied by changes in their status, including increased education, occupation, income, and, especially, relative contribution to the overall family income (Blau and Ferber 1986).

Some of the improvements in women's economic situation have resulted from broad technological, demographic, and economic shifts such as declining family size, later age at marriage, improved means of birth control, increasing inflation (which makes it necessary to have two paychecks to support a family), and expanding employment in the traditionally female service and clerical sectors. Many of these gains could not have been won without collective struggle in the feminist movement. Regardless of the cause, to the extent that experience in the workplace conditions class identification, women should draw more upon their own experiences, and less on their husbands', in assessing where they stand in the class structure. Clearly, the attitudes of both

men and women toward women working outside the home and more equal pay for equal work have also become more favorable in the past three decades (see, e.g., Mason et al. 1976; Cherlin and Walters 1981). To the extent that these behavioral and attitudinal changes have taken place, wives may now attach more importance to their own contribution to their class standing. At the same time, husbands may now attribute more importance to their wives' contribution to the family's class position.

In this paper, we employ data from the National Opinion Research Center's General Social Survey to analyze the class identification of married men and women during the 1970s (1974 to 1978) and 1980s (1980 to 1985). All of the previous research cited above was based on data collected during the early to mid-1970s and, as we argued above, we believe that the class identification of both husbands and wives may have since changed.

In our analysis, we regress the individual's perception of his or her class position on indicators of the individual's own objective position in the stratification system and on comparable indicators of the spouse's position. The most complete previous study (Jackman and Jackman 1983, p. 146) analyzed the effect of both the individual's and spouse's education, occupational status, and income. Our analysis also includes the individual's and spouse's ownership of a business as an indicator of their class position. In classical and contemporary class theories, ownership and authority relations have been argued to constitute the fundamental bases of class relations (see, e.g., Wright 1976; Robinson and Kelley 1979). These variables have also been found to influence individuals' class identifications (Coleman and Rainwater 1978; Robinson and Kelley 1979; Wallace and Jepperson 1986). While we have a measure of business ownership for both individuals and their spouses, we do not have a measure of authority for spouses. Were we to include the individual's position in authority relations without controlling for spouse's position, we would be unable to judge the extent to which individuals take into account their own characteristics as opposed to their spouses'. Thus, we regress the individual's class identification on his or her education, occupational prestige, business ownership, and income, as well as on the same characteristics for the spouse. Additional analyses including the authority variable for individuals yielded essentially the same substantive conclusions as reported below (details are available on request from the authors).

We examine three models of class identification for married men and women:

a. In the *independence* model, individuals rely on their own characteristics and not their spouses' in determining their class position.
b. In the *sharing* model, equal weight is attributed to both spouses' characteristics.
c. In the *borrowing* model, only the spouse's characteristics are taken into account.[1]

We view these models as a continuum ranging from treatment of one's own characteristics as the sole determinant of one's class position ("independence"), through treatment of them as joint determinants along with the spouse's characteristics ("sharing"), to treatment of them as unimportant relative to the spouse's characteristics ("borrowing").

We employ data for the 1970s and 1980s to address the following questions:

a. Are married women increasingly taking more account of their own characteristics (and, a separate question, less account of their husbands' characteristics) in determining their class position? In other words: Is the model for wives moving from a "borrowing" model to a "sharing" or even an "independence" one?
b. Are married men beginning to take into account their wives' characteristics (and perhaps, decreasingly taking into account their own characteristics) in determining their class position? In other words: Is the model for husbands moving from an "independence" model to a "sharing" model?
c. Is the process of class identification becoming more similar for husbands and wives? Are they both moving toward a "sharing" or an "independence" model?

DATA AND METHODS

Samples

The data are from the National Opinion Research Center's General Social Surveys for 1974 to 1985 (NORC 1985).[2] These are national samples of the noninstitutionalized population of the United States, 18 years of age or older. To ensure that all respondents are of working age, we placed an upper limit on age of 64. Because of sample restrictions necessary in computing spouse's income (see below), there is an insufficient number of cases in many of our subsamples for each survey year to allow analysis of these separately. We have, therefore, pooled samples into one sample for the 1970s (five samples from 1974 to 1978) and one for the 1980s (five samples from 1980 and 1982 through 1985).[3]

We analyze three subsamples for the two decades: (1) employed married men with nonemployed wives and nonemployed married women with employed husbands ($N=590$ men and 750 women in the 1970s, and 363 men and 373 women in the 1980s); (2) employed married men with employed wives and employed women with employed husbands ($N=565$ men and 647 women in the 1970s, and 652 men and 738 women in the 1980s); and (3) single men and women employed full-time ($N=203$ men and 159 women in the 1970s and 244 men and 189 women in the 1980s).

Measures

Class identification is respondent's self-placement in a choice of four classes: lower (scored 1), working (2), middle (3), and upper (4), in response to the question: "If you were asked to use one of four names for your social class, which would you say you belong in?" Respondent's and spouse's *education* is measured as years of schooling. Respondent's and spouse's *occupational prestige* is a scale (ranging from 9 to 82) developed by Robert W. Hodge, Paul S. Siegel, and Peter H. Rossi for NORC (see Siegel 1971). Respondent's and spouse's *ownership* is indicated by whether they owned a business (coded 1) or not (coded 0). *Respondent's income* in the year prior to the survey (in thousands of dollars) is measured in 13 categories for 1974, 1975, and 1976, in 17 categories for 1977, 1978, and 1980, and in 18 categories for 1982 through 1985, with midpoints representing each category. To adjust for inflation, we used the Consumer Price Index (U.S. Department of Commerce 1985) to

---

[1] When they found that both husbands and wives tended to use the husband's characteristics in determining the family's class position, Jackman and Jackman (1983, p. 140) referred to this as a "borrowing" mode. We prefer to label this an "independence" model for the husband and a "borrowing" model for the wife. This permits the conceptualization of the three models as a continuum and allows for the possibility that husbands and wives may hold different models of class identification.

[2] Since respondent's income was not asked in 1972 and 1973, these survey years have been dropped from the analysis. The GSS was not conducted in 1979 and 1981.

[3] The number of cases in each survey year was too small to yield stable coefficients from year to year. As a check on the validity of our conclusions, we pooled successive groups of three consecutive survey years (i.e., 1974, 1975 and 1976; 1975, 1976, and 1977; etc.), in effect computing a "moving average" of the regression coefficients. These analyses yielded essentially the same substantive conclusions as we report below for the pooled samples for the 1970s and 1980s.

Table 1. Class Identifications of Married Men and Women, Age 18 to 64, By Employment Status of Spouse

Class Identification	Men		Women	
	1974–1978	1980–1985	1974–1978	1980–1985
Only husband employed ·				
Upper	2%	5%	3%	3%
Middle	44	41	51 ·	47
Working	52	53	43	46
Lower	2	2	3	5
Total	100%	100%	100%	101%
N of cases	(590)	(363)	(750)	(373)
Both spouses employed				
Upper	3%	3%	2%	2%
Middle	45	44	45	44
Working	51	51	52	53
Lower	1	2	1	1
Total	100%	100%	100%	100%
N of cases	(565)	(652)	(647)	(738)

convert the income measure for each year into 1984 dollars. While we have no direct measure of *spouse's income,* we are able .to create a reasonable proxy for this. The NORC data include questions (referring to the previous year) on respondent's income, family income, the number of earners in the household, and the current employment status of both spouses. To compute spouse's income, we subtracted the respondent's income from family income. To ensure that this produces a reasonable estimate of·spouse's income, we restricted our analysis to households where both spouses were currently employed (including temporarily laid off or ill), where there were only two earners in the previous year,[4] and ·where the respondent's income was not zero.[5]

[4] In the 1970s sample, for employed respondents with employed spouses, 72 percent of men and 72 percent of women had two earners in the household the previous year. The figures for the 1980s sample are 76 percent and 74 percent, respectively.

[5] There are two possible sources of error in estimating spouse's income in this way. First, since employment status of the husband and wife is for the current year but the income data refer to the previous year, it is possible that one of the two earners was someone other than the husband or wife, e.g., a child. Our restriction to couples who are both currently employed and the further condition that the respondent's income not be zero (so that we know one of the two earners was the respondent) lessens the possibility that this source of error may arise. Second, since family income includes income from all sources and not just job income, some of the income attributed to spouses may arise from a couple's nonjob sources of income. We believe that errors from this source are relatively slight. While we have no estimate of the extent of nonjob income for the samples we analyze, in couples where the husband was the only earner, family income from all sources exceeded male respondent's job income by only 5.6 percent and 4.6 percent, respectively, in the 1970s and 1980s data, and the correlation between

For couples in which the wife was not currently employed, we further restricted the sample to households where there was only one earner in the previous year. When the respondent in such couples was a woman, husband's income was estimated from family income. When the respondent was a man, we required that there be only one earner in the previous year and that the respondent's income in the previous year be greater than zero (to be certain that it was the husband who had earned the income in that year).[6] The effect of our sample restrictions is to limit analysis to couples that had the same work arrangement in the household for the survey year and the previous year (i.e., only husband employed or both spouses employed).

We estimate our models by ordinary least squares (OLS) regression methods. Since most of the cases on the dependent variable, class identification, fall into the middle two categories (see Table 1), it could be argued that (after collapsing to two categories) a logit analysis would be more appropriate. In analyses such as ours, where the split on the dependent variable is not highly skewed, however, logit and OLS regression analyses yield essentially the same results (Hanuskek and Jackson 1977, p. 185).

RESULTS

Class Identification

The distributions of class identification for married people are rather similar across genders and over time (see Table 1). Most respondents place themselves in either the middle or working class, with slightly more choosing the latter

respondent's income and family income was .94 and .95, respectively..

[6] Unfortunately, there were too few cases of couples where only the wife was employed to permit analysis of this interesting subsample.

category. The only exceptions to the general pattern occur among nonemployed wives. In the 1970s, these women were slightly more likely than other categories of women and men to see themselves as middle class. By the 1980s, however, their class perceptions have come to be somewhat more in line with other sex/employment categories.

Couples with a Nonemployed Wife

One would expect that the most traditional couples—i.e., those in which both spouses accord the husband's characteristics the greatest weight and the wife's the least weight in determining their class position—would be those in which the wife is not employed. Since nonemployed wives do not have a paid occupation or income, their education is the only characteristic that they or their husbands might take into account in determining class placement. Table 2 confirms the expectation that such couples attach the greatest importance to the husbands' characteristics. In the 1970s, husbands took into account their own prestige and income and their wives' education in assigning themselves to a class. Analysis of the variance explained by the husband's own characteristics alone, by his wife's characteristics alone, and by both sets of characteristics suggests that husbands are not attaching much weight to their wives' characteristics in determining their class position. For these employed men, 28 percent of the variance is explained by both their own and their spouses' characteristics (see Table 2). Wife's education adds less than 1 percent to the variance explained by husband's education, prestige, ownership, and income (27 percent). Reversing the procedure, we find that the husband's characteristics add 17 percent to the 11 percent of the variance explained by wife's education. Thus, while these husbands in the 1970s attached some importance to their

spouses' education, they apparently saw their own occupational prestige and income as far more salient to their class position.

In the 1980s, men with nonemployed wives moved toward an even more pronounced independence model. Wives' education is now unimportant to these men's class placement (the decline in slope is significant at the .05 level), and their own education is now significant, in addition to their occupational prestige and income. It is possible that this reflects a decline in the status of nonemployed women as increasing numbers of women enter the labor force and as an employed wife becomes the norm. Regardless of the cause, husbands of nonemployed wives moved toward a more pronounced independence model.

Nonemployed wives in the 1970s exhibited a borrowing model. Their own education was insignificant and their husbands' education, occupational prestige, business ownership, and income were significant. In the 1980s, nonemployed married women are now taking into account their own education (although the change in slope is not significant) as well as their husbands' income, education, occupational prestige, and ownership. These women have shifted slightly in the direction of a sharing model, in which their own characteristics are accorded at least some modest importance. Even in the 1980s, however, nonemployed wives, like their husbands, attach far more weight to their husbands' than to their own characteristics in determining their class position. As we discussed in the introduction, while couples such as these, in which the wife is not employed, used to be the norm, the increasingly more common situation is that to which we now turn—couples with both spouses employed.

Both Spouses Employed

As we noted above, research during the 1970s on class identification among couples with both

Table 2. Determinants of Subjective Class Identification Among Employed Men, Age 18 to 64, with Nonemployed Wives and Among Nonemployed Women, Age 18 to 64, with Employed Husbands

Variables	Men		Women	
	1974–1978	1980–1985	1974–1978	1980–1985
Education	.013	.030*	.016	.024*
Prestige	.007***	.008**		
Ownership	.057	.054		
Income	.010***	.011***		
Spouse's Education	.025*	−.004	.034***	.030*
Spouse's Prestige			.007***	.005*
Spouse's Ownership			.170**	.162*
Spouse's Income			.013***	.012***
R^2	.278	.292	.248	.325
N	540	318	683	329

* Regression coefficient significant at $p<.05$, one-tailed test.
** Regression coefficient significant at $p<.01$, one-tailed test.
*** Regression coefficient significant at $p<.001$, one-tailed test.

Table 3. Determinants of Subjective Class Identification Among Full-time Employed Respondents, Age 18 to 64, with Full-time Employed Spouses

	Men		Women	
Variables	1974–1978	1980–1985	1974–1978	1980–1985
Education	.043**	.038**	.020	.045**
Prestige	.006*	.004*	.003	−.003
Ownership	.111	.053	.090	.108
Income	.005*	.014***	.003	.010*
Spouse's Education	−.017	.016	.037***	.012
Spouse's Prestige	.002	.001	.000	.003
Spouse's Ownership	.044	−.046	.171*	.018
Spouse's Income	.007*	.005	.007***	.009***
R^2	.177	.318	.221	.178
N	293	336	375	374

* Regression coefficient significant at $p<.05$, one-tailed test.
** Regression coefficient significant at $p<.01$, one-tailed test.
*** Regression coefficient significant at $p<.001$, one-tailed test.

spouses employed found that husbands attached greater importance to their own than to their wives' characteristics, but arrived at inconsistent findings with regard to wives. The study that is most comparable to ours and one that overcame the mis-specification that plagued some earlier analyses is Jackman and Jackman (1983). They found that in 1975, the one exception to the tendency for both sexes to take into account only the husband's characteristics was for employed wives to consider their own education—an attribute that they argue is largely fixed before marriage (but see Davis and Bumpass 1976) and not linked to experience in the labor force.

Table 3 presents results for men and women who are employed full-time and whose spouses are employed full-time.[7] Our findings for the 1970s generally parallel Jackman and Jackman's (1983, p. 145). In both the 1970s and 1980s, husbands attach far more weight to their own characteristics than to their wives'. The husband's own education, occupational prestige, and income are significant in both decades. Husbands in the 1970s also considered their wives' income. The overall pattern in the 1970s, however, is for a husband's characteristics to greatly outweigh his spouse's in determining his class position.

By the 1980s, married men take into account their own education, occupational prestige, and income but no longer consider their wives'

---

[7] We have restricted this analysis to *full-time* employed respondents and spouses, since the greater proportion of women than men who work part-time (Blau and Ferber 1986, p. 128) and the perhaps weaker commitment of part-time workers to the labor force might confound male/female differences in class identification. In fact, essentially the same substantive conclusions are reached when all (full-time and part-time) employed men and women are analyzed (details available on request from the authors).

income. While the decline in the slope of wife's income from the 1970s to the 1980s is not significant, the increase in the slope of the husband's own income is significant ($p < .001$). Thus, as we found for the husbands of nonemployed wives, whatever movement among married men from the 1970s to the 1980s was toward an even more pronounced independence model.

The results for full-time employed women suggest a shift away from a borrowing model to a sharing one. In the 1970s, only their husbands' characteristics (education, business ownership, and income) were significant. In fact, in the 1970s, employed women differed little from their nonemployed counterparts in their tendency to attach almost no importance to their own characteristics in determining their class position. In the 1980s, however, the pattern is quite different. Employed wives now take into account their own education and income, and they no longer take into account their husbands' education or ownership. While the increases in women's own education and income just fail to be significant at the .05 level, the declines in their husbands' education and ownership are significant at that level.

Although arriving at different end-points, both married men and women may be seen as moving toward models that attach greater importance to the individual's own characteristics. For married men, the shift is toward a slightly more pronounced independence model. For married women, it is away from a borrowing model, which accords little import to the wife's characteristics, to a sharing model, which acknowledges the contributions of both spouses.

Status Maximization?

We have seen that even men with full-time employed wives take into account their own

Table 4. Determinants of Subjective Class Identification Among Full-time Employed Single Respondents, Age 18 to 64

Variables	Men		Women	
	1974–1978	1980–1985	1974–1978	1980–1985
Education	.061***	.065***	.028	.078***
Prestige	.006*	.004	−.002	.001
Ownership	.011	.302**	.249	.111
Income	.003	.008**	.011*	.004
Father's Education	.010	.015	.028*	−.006
Father's Prestige	.005	.001	.006	.000
Father's Ownership	.173*	.050	−.100	.129
Mother's Education	.014	−.029	.023	.017
Mother's Employment	.015	.053	−.088	−.115
R^2	.278	.273	.223	.157
N	203	244	159	189

    * Regression coefficient significant at $p<.05$, one-tailed test.
   ** Regression coefficient significant at $p<.01$, one-tailed test.
  *** Regression coefficient significant at $p<.001$, one-tailed test.

characteristics more than their wives'. It may be that these husbands are attaching greater weight to their own characteristics, as opposed to their wives', only to justify the highest possible class placement. Coleman and Rainwater (1978, p. 43) found that individuals place themselves in the highest class that they feel can be justified.

In analyses not reported here (available on request from the authors), we examined couples in which the wife's education, occupational prestige, or income is higher than her husband's in order to see if husbands and wives accord greater weight to the wife's characteristics under such circumstances.[8] With the exception of husbands in the 1970s who earned less than their wives (15 percent of cases), there was no evidence of "status maximization." In general, couples in which the wife's education, prestige, or income was higher than the husband's showed no tendency to attach greater importance to the wife's characteristics than did the couples we have discussed above.

Single Men and Women

We have seen above that women and men attach different weights to their own and their spouses' characteristics in determining their class position. Men consistently place great importance on their own characteristics, while women either attach little importance to their own characteristics or, at best, accord equal importance to their own and their husbands' characteristics. In order to see whether, even apart from marriage, men and women have different ways of assigning themselves to classes, we present models in Table 4 explaining class identification for single

(never married) men and women who are employed full-time. Since these people tend to be young, we include in the model, in addition to the four respondents' characteristics we have previously analyzed, five attributes of parents: father's education; mother's education; father's occupational prestige (when the respondent was 16); father's business ownership (when the respondent was 16); and mother's employment status (whether she was employed for as long as a year after she was married).[9] In addition to allowing us to examine class identification among women and men apart from marriage, these models have the advantage of affording an analysis of whether women and men have different patterns of borrowing status from their parents, sharing status with them, or holding an independence model, in which only their own, and not their parents', characteristics matter in their class placement.

In the 1970s, both single men and women held a sharing model in which both their parents' (especially their fathers') and their own characteristics played a role in determining their class identification (see Table 4). Single women in the 1970s clearly had more of a tendency to borrow from their parents, however. Their own education, prestige, ownership, and income added 4 percent to the 18 percent of the variance explained by family characteristics, while single men's own characteristics added 16 percent to the 12 percent explained by their parents' characteristics. Single men in the 1970s took into account their fathers' business ownership, their own prestige, and, especially, their own education in determining their class position.

---

[8] Unfortunately, there was an insufficient number of wives who were higher than their husbands on all three characteristics to consider the effect of all characteristics together.

[9] There is no measure of parents' income comparable to that available for respondents, nor are there measures of mother's occupational prestige or mother's business ownership.

Single women in the 1970s took into account their fathers' education and their own income.

In the 1980s, single men are attaching much less importance to their parents' characteristics than did their counterparts in the 1970s. The variance explained by family characteristics alone declined from 12 percent to 4 percent, while the overall variance explained by both parents' and own characteristics remained virtually unchanged (28 percent in the 1970s versus 27 percent in the 1980s). In the 1980s, no family characteristic is significant for single men, and their own education, ownership, and income are significant. Thus, single men have moved from a sharing model to an independence model.

Single women in the 1980s also moved toward an independence model; the variance explained by parents' characteristics declined dramatically from 18 percent to 4 percent. This was accompanied by a decline in the overall variance explained by both parents' and the respondent's own characteristics from 22 percent to 16 percent. In the 1980s, single women take into account only their own education in placing themselves into a class.

Thus, both single men and women moved toward an independence model and away from a sharing one. This finding is entirely consistent with the tendencies we found above for married men and women. Married men are increasingly taking into account their own, as opposed to their spouses', characteristics. Married women too are moving in the direction of counting their own contribution to their class position, though movement for them is from a borrowing to a sharing model, not to a full independence model.

SUMMARY AND DISCUSSION

We began our paper expecting that the class identifications of men and women might have changed considerably over the past decade. We hypothesized that couples with both spouses employed may have moved away from a model in which only the husband's characteristics are salient to class identification toward a sharing model in which the contributions of both spouses are acknowledged. While we have found a tendency for both employed and (to a lesser extent) nonemployed wives to take more account of their own characteristics and move toward a sharing model, we have also found some tendency for husbands to take *less* account of their wives' characteristics and more account of their own.

Among women, we may be witnessing a larger change in how they assess who they are, including their class identity. This may be

occurring because more women have direct connections to the labor market and can use their own objective characteristics to place themselves in a class. It may also be the result of a change over time in the salience of employment for employed women. When only a minority of women were involved in the relations of production, the general cultural consensus was that family status, not employment position, was the salient identity for women. In such circumstances, even employed women may have seen their identity as family-based, with their class position derived from the husbands and fathers with whom they lived. While women's economic position is still very much conditioned by the presence and earnings of men in the household, employment has become the statistical norm for both married and single women. Thus, women's class identity may be changing in the 1980s, not only because there are more women with direct connections to the relations of production, but also because the meaning of such a connection has taken on more salience for female identity.

While it is tempting to interpret the shift among men toward a more pronounced independence model as some sort of antifeminist backlash against women's changing role in the family and labor force, other findings in our analyses lead us to suspect that something else may be at work here—something that may play a role in the changing class identifications of both men and women. We conceptualized the three models of class identification as a continuum from "borrowing" to "sharing" to "independence." Viewed in this sense, both husbands and wives are moving in the direction of increasing independence. The husbands, who already were rather independent, attribute even more importance to their own characteristics and less importance to their (employed or nonemployed) wives'. The wives, who held a borrowing model in the 1970s, moved toward a model in which their own characteristics mattered more and their husbands' characteristics less.

Bellah and his coauthors (1985) have documented a tendency toward increased individualism in American life. They observe that "marriage and the family, while still desirable, are now in several ways optional" (1985, p. 110). Even among those who do marry, they find a tendency toward increasing individualism or independence of partners from each other. Similarly, Ehrenreich (1983) describes a shift over the past 30 years in popular culture, medicine, and psychology that increasingly emphasizes the desirability of men's independence from women and children.

The increasing individualism that permeates American life may thus be partly responsible for

the movement of both husbands and wives toward an independence model. Were there no further evidence in our data of this tendency, we would feel less confident in making this speculation. But we found much the same pattern for single men and women. In the last decade, both single men and women have moved from a sharing model, in which *both* their own *and* their parents' (especially their fathers') characteristics were important, to an independence model, in which only their own matter. As in our analyses of married people, the most striking changes were among women, suggesting that increased individualism may have gone hand in hand with fundamental changes in attitudes about women's position in society and about the feasibility of women having an identity of their own, apart from their fathers and husbands.

In the case of women, it is important to note that their increasing individualism and recognition of their own contribution resulted partly from several generations of *collective* effort that empowered women. The feminist movement, together with broad economic and demographic shifts (smaller family sizes, later ages at marriage, higher divorce rates, expansion of the female-dominated service and clerical sectors, high inflation, etc.), set the stage for dramatic increases in the number of women who are no longer confined only to unpaid domestic labor but also are directly participating in class relations. The political mobilization of women helped change not only the objective position of women but also their awareness of the contribution made by women, both employed and nonemployed, to their family's economic position. In this regard, changes in class identification—i.e., the shift away from male-derived identities—may signal new configurations of power within families as women press for more equal partnership in family decision-making, the disposition of family resources, and the division of domestic labor.

In conclusion, our findings suggest that such changes as the increasing participation of women in the labor force, declining family size, the growing number of women who live in households without men, and changing attitudes about sexual divisions of labor and power in the family and workplace require reexamination of the assumption of many theorists that only the husband's labor force characteristics should be taken into account in placing a family in the stratification system. This assumption is increasingly incongruent with the reality of work arrangements in contemporary American marriages and, as we have seen, would not find much support among the growing segment of American women who work in the labor force.

REFERENCES

Acker, Joan. 1973. "Women and Social Stratification: A Case of Intellectual Sexism." *American Journal of Sociology* 78:936–45.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart: Individualism and Commitment in American Life*. New York: Harper and Row.

Blau, Francine D. and Marianne A. Ferber. 1986. *The Economics of Women, Men, and Work*. Englewood Cliffs, NJ: Prentice-Hall.

Britten, Nicky and Anthony Heath. 1983. "Women, Men, and Social Class." In *Gender, Class, and Work*, edited by Eva Gamarnikow et al. London: Heinemann.

Cherlin, Andrew and Pamela B. Walters. 1981. "Trends in United States Men's and Women's Sex Role Attitudes: 1972–1978." *American Sociological Review* 46:453–60.

Coleman, Richard P. and Lee Rainwater. 1978. *Social Standing in America: New Dimensions of Class*. New York: Basic Books.

Dahrendorf, Ralf. 1959. *Class and Class Conflict in Industrial Society*. Stanford: Stanford University Press.

Davis, Nancy J. and Larry L. Bumpass. 1976. "The Continuation of Education after Marriage among Women in the United States: 1970." *Demography* 13:161–74.

DeFronzo, J. 1981a. "Class Identification among Working Wives." *Psychological Reports* 49:171–77.

———. 1981b. "Differences in the Determinants of Husbands' and Working Wives' Class and Political Party Identification" *Sex Roles* 7:355–61.

Ehrenreich, Barbara. 1983. *The Hearts of Men*. Garden City, NY: Anchor/Doubleday.

Felson, Marcus and David Knoke. 1974. "Social Status and the Married Woman." *Journal of Marriage and the Family* 36:516–21.

Giddens, Anthony. 1973. *The Class Structure of the Advanced Societies*. London: Hutchinson.

Goldthorpe, John H. 1983. "Women and Class Analysis: In Defense of the Conventional View." *Sociology* 17:465–88.

Hanushek, Eric A. and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic.

Hiller, Dana V. and William W. Philliber. 1978. "The Derivation of Status Benefits from Occupational Attainments of Working Wives." *Journal of Marriage and the Family* 40:63–70.

Jackman, Mary R. and Robert W. Jackman. 1983. *Class Awareness in the United States*. Berkeley: University of California Press.

Mannheim, Karl. 1936. *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York: Harcourt, Brace and Company.

Marx, Karl. [1845–46] 1978. *The German Ideology*, Part I. Pp. 146–200 in *The Marx-Engels Reader*, 2nd ed., edited by Robert C. Tucker. New York: Norton.

———. [1893] 1909. *Capital*, Vol. 3. Translated from the first German edition by Ernest Untermann. Chicago: Charles H. Kerr & Co.

Mason, Karen Oppenheim, John L. Czajka, and Sara Arber. 1976. "Change in U.S. Women's Sex-Role Attitudes, 1964–1973." *American Sociological Review* 41:573–96.

National Opinion Research Center. 1985. *General Social*

Surveys, 1972–1985: Cumulative Data [MDRF]. Chicago: National Opinion Research Center [producer]. New Haven: Roper Public Opinion Research Center, Yale University [distributor].

Parkin, Frank. 1971. *Class Inequality and Political Order*. London: MacGibbon and Kee.

Philliber, William W. and Dana V. Hiller. 1978. "The Implication of Wife's Occupational Attainment for Husband's Class Identification." *Sociological Quarterly* 19:450–58.

————. 1979. "A Research Note: Occupational Attainments and Perceptions of Status among Working Wives." *Journal of Marriage and the Family* 41:59–62.

Ritter, Kathleen and Lowell Hargens. 1975. "Occupational Positions and Class Identifications of Married Working Women: A Test of the Asymmetry Hypothesis." *American Journal of Sociology* 80:934–47.

Robinson, Robert V. and Jonathan Kelley. 1979. "Class as Conceived by Marx and Dahrendorf: Effects on Income Inequality and Politics in the United States and Great Britain." *American Sociological Review* 44:38–58.

Rossi, Peter, H.W.A. Sampson, C.E. Bose, G. Jasso, and J. Passel. 1974. "Measuring Household Social Standing." *Social Science Research* 3:169–90.

Siegel, Paul S. 1971. *Prestige in the American Occupational Structure*. Unpublished Ph.D. diss., University of Chicago.

U.S. Bureau of the Census. 1975. *Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition, Part 1*. Washington, DC: Government Printing Office.

————. 1985. *Statistical Abstract of the United States: 1986*, 106th Ed. Washington, DC: Government Printing Office.

U.S. Department of Commerce. 1985. *Business Statistics, 1984*. Washington, DC: Government Printing Office.

Van Velsor, Ellen and Leonard Beeghley. 1979. "The Process of Class Identification among Employed Married Women: A Replication and Reanalysis." *Journal of Marriage and the Family* 41:771–78.

Wallace, Michael and Ronald L. Jepperson. 1986. "Class Structure and Political Culture: Evaluations of Key Political Groups in Eight Western Capitalist Nations." Pp. 321–361 in *Research in Social Stratification and Mobility*, Vol. 5, edited by R.V. Robinson. Greenwich, CT: JAI Press.

Weber, Max. [1922] 1969. "Class, Status, and Party." Pp. 24–34 in *Structured Social Inequality*, edited by Celia S. Heller. London: Macmillan.

Wright, Erik Olin. 1976. "Class Boundaries in Advanced Capitalist Societies." *New Left Review* 98:3–41.

ASPECTS OF MIGRATION IN AN ADVANCED INDUSTRIAL SOCIETY*

FRANKLIN D. WILSON

University of Wisconsin-Madison, Wisconsin

This paper tests the hypothesis that 1935–80 U.S. migration patterns both within and between metropolitan and nonmetropolitan areas and regions and migrant/nonmigrant educational differences reflect regional changes in socioeconomic development and settlement patterns. Gross migration flows support the mobility transition hypothesis and the idea that developmental changes in metropolitanization structure the context of migration. Since the Great Depression, socioeconomic transformations of the periphery have reduced differences in migration patterns between the periphery and core regions. While migrants still have more education than nonmigrants, their differences vary systematically by regional socioeconomic development. Specifically, migrant/nonmigrant educational differences are (1) similar at origin and destination when both are at similar levels of socioeconomic development; (2) lower at the origin if the origin is higher in socioeconomic development; and (3) higher at the origin if the destination is higher in socioeconomic development.

INTRODUCTION

After World War II, the United States moved into a postindustrial stage of societal development. This involved a shift from heavy manufacturing to tertiary and high technology industries; technological changes in the production and distribution of goods and in the processing and transmission of information; and the emergence of multinational corporations as the major stimulus of employment growth, innovation, investment, and spatial relocational activities (Noyelle and Stanback 1985 pp. 17–25). Scholars have shown how the postindustrial order altered the spatial structures of communities and patterns of population redistribution (e.g., Goldscheider 1983; Frey 1987; Long 1985). They revealed an association between societal development, urbanization/metropolitanization, and migration. Thus, the rate,

direction, and character of migration flows reflected the stage of societal development and its system of urban agglomeration. Long (1985), for example, suggested that the movement toward higher levels of socioeconomic development stimulated migration, which, in turn, changed the pattern of spatial settlements in the nation.

This paper focuses on migration both as an effect of changing socioeconomic conditions and as a contributor to regional convergence in socioeconomic structure. First, I ask whether migration flows by settlement pattern, within and between regions, change with changes in socioeconomic conditions. Second, I test whether area differences in socioeconomic and opportunity structures lead to (1) the generation of migration streams that differ in size and composition; and (2) human capital differences between migrant and nonmigrants (see Lee 1966; Campbell and Johnson 1976). The central question is whether the migration patterns of an advanced industrial society reflect a process of regional convergence in which the benefits of socioeconomic development are spread evenly throughout the society.

THEORETICAL BACKGROUND

Zelinsky's (1971) model of mobility transition is used to develop expectations regarding the timing and extent of change in migration patterns that connect functionally distinct geographic areas. He observes several associations between migration and modernization. First, modernization leads to a series of population-related transitions, including the well-known demographic transition in vital rates and spatial

mobility. Critically, the form and intensity of migration changes as societal modernization moves toward more advanced stages of socioeconomic development. Changes follow in the function, frequency, direction, periodicity, distance, and composition of migration streams. Second, although modernization promotes spatial differentials in economic opportunities within early stages, they diminish in the advanced stages. Transition to advanced stages also increases migration, because it increases the range of locations in which individuals can seek personal fulfillment. At the highest stage of development, a population equilibrium is reached, as well as a convergence in the social and economic structure of metropolitan and nonmetropolitan areas. Here, migration becomes essentially a random process, no longer substantially affected by area differences in resources, labor, or capital.

Zelinsky observes that, since the 1930s, the United States has been in the "advanced societal" stage of development. This stage is characterized by (1) a leveling off of migration at a higher level; (2) continued movement from countryside to city, but at a decreasing rate in both absolute and relative terms; (3) vigorous movement of migrants between and within urban agglomerations; and (4) the emergence of noneconomic motivations as a significant factor in migration. A number of other studies support findings consistent with some of Zelinsky's observations on U.S. redistribution trends (for reviews see Brown and Sanders 1981; Wilson 1984, 1987; Fuguitt 1985).

One major shortcoming of equilibrium and convergence models of urbanization and associated patterns of migration is that these models rarely consider regional differences in the timing and extent of socioeconomic development. For example, several studies indicate that the initial stage of industrialization in currently developed countries was characterized by polarized spatial development where labor and capital became concentrated in a core region (e.g., Morrill 1980; Long 1981; Freidmann 1973; Vining 1986). As development proceeded, this pattern of urban/regional concentration gave way to a more decentralized and deconcentrated urban pattern within the core region and migration from the core to the developing periphery.

The substantial role of U.S. regional population redistribution in the metropolitanization process indicates that Zelinsky's mobility transition hypothesis needs to be modified to take account of this source of variation. Available evidence indicates that the U.S. has not yet reached the stage where the spatial distribution of its population is at an equilibrium point. The general hypothesis of uneven regional development, which considers the relation between the

core and the periphery, applies to the historical and contemporary development of the U.S. urban system (see Wilson 1986, 1987).

Several studies reveal that the spatial distribution of population changes as technological and structural forces alter product demand, labor/capital cost ratios, and the relative advantages of geographical areas with respect to their ability to provide the appropriate mix of resources to promote productivity, efficiency, and profitability.[1] The apparent rise in the economic fortunes of the Sunbelt or periphery region of the United States results directly from these changes. These changes have also altered the value and character of interregional and metropolitan and nonmetropolitan migration flows. The primary objectives of this paper are to describe changes in these patterns of flows and assess their impact on the changing spatial structure of the U.S. regional system.

As noted previously, one of the primary goals of this paper is to determine whether migration flows, by settlement pattern, within and between regions, change with changes in socioeconomic condition. In focusing on change in migration flows it is particularly important to ascertain whether the pattern of change is leading to regional convergence in settlement pattern and migration flows between settlements. Figure 1 graphically portrays the hypothesized dominant direction of migration flows connecting metropolitan and nonmetropolitan areas within and between core and periphery regions. The definitions of these regions are adapted from Freidmann (1973) and Morrill (1980). Core regions have typically achieved an advanced stage of industrialization, and high concentration of control, coordination, and decision-making functions that direct the flow of information, goods, and people to and from the

---

[1] Smith (1984, p. 115), for example, notes that, historically, regional differences in resource endowment were decisive in promoting regional differences in development and metropolitanization, but that, currently, uneven development is based more and more systematically upon spatial differentiation of the social determinants of capital accumulation such as wage rates, level of unionization, and workers' skill and productivity. Other writers tend to emphasize the substantial impact of shifting political alliances and the military-industrial complex on contemporary regional differences in population and economic growth (Abbott 1981; Mollenkopf 1983); the spatial impact of product life cycles, and how the changing demands of capitalization produce a dynamic disequilibrium with spatial implications (see Frey 1987; and Weinstein et al. 1985 for reviews); and technologically induced changes in production systems, which lead to locational changes as the latter takes into account spatial variations in the cost of labor and capital (see Norcliffe 1984; Gertler 1986; Scott 1986, for reviews).

'periphery, as well as a high degree of metropolitanization, with intermetropolitan movements being the dominant form. This description has long characterized the Northeast, North Central, and the Pacific (mainly California) regions. Although the periphery, because of the heavy emphasis on extractive activities, has been economically subordinate to the core since the 1940s, the periphery's economy has grown rapidly, based on its greater ability to innovate and perform seedbed functions for emerging industries. Its rate of metropolitanization has also increased because of growing migration from the core and nonmetropolitan areas within the region. Post-Depression changes in the South and Mountain regions fit this descripion.

Historical differences in the timing and extent of socioeconomic development and metropolitanization between the core and periphery should result in different patterns of migration within and between these regions. The hypothesized rankings of the dominant migration streams within and between sectors in each cell of Figure 1 apply to the initial period of 1935–40. First, advanced socioeconomic development and metropolitanization in the core promotes the dominance of metropolitan origin streams to both intraregional and interregional destinations. The intermetropolitan flows will be the dominant stream originating from the core because metropolitan areas contain the majority of the population and they dominate the origin and destination of social and economic transactions. The metropolitan to nonmetropolitan stream is the second largest, but the reasons are different. Here, the size of the within-core stream reflects individual metropolitan areas reaching advanced stages of population deconcentration. As metropolitan areas age, popula-

tion deconcentration first appears as suburban movements and later as central city and suburban movements to adjacent nonmetropolitan areas (Norcliffe 1984; Long 1985; Scott. 1986; Wilson 1986). The size of the core to periphery, metropolitan to nonmetropolitan stream, on the other hand, is probably more associated with the formation of metropolitan areas in the periphery than with higher levels of deconcentration in existing metropolitan areas (see Wilson 1986).

The relative size of the migration streams originating from the periphery should differ from streams originating from the core region. The principle feature of the periphery is that its subordinate status is being transformed as a result of social and economic changes. Moreover, the fact that a majority of the periphery's population was concentrated in nonmetropolitan areas prior to the 1960s indicates that migration from it would be larger than from the metropolitan sector. Whether the nonmetropolitan to metropolitan stream is larger than the inter-nonmetropolitan one within the periphery depends on whether opportunities for employment are increasing faster in metropolitan than in nonmetropolitan areas. A good indicator of whether, initially, opportunities are greater in the core is whether migrants from the periphery migrate more to metropolises in the core. The hypothetical ranking in the lower left-hand cell of Figure 1 predicts that this pattern will prevail.

Changes in the relative ranking of the within- and between-regional metropolitan and nonmetropolitan streams are of greater interest, since they should reflect the changing socioeconomic standing of the core and periphery prior to and during the initial stage of the periphery's development. Initially, the dominant flow of interregional migration will be from the periphery to the core, followed by a reversal. This reversal corresponds to changes in the relative growth of employment opportunities, which will be higher in the periphery over time. Increased regional similarity in socioeconomic structures could also increase interregional migration (see Wardwell 1980). In addition to these general interregional migration changes, the following are also expected: (1) the pattern of movement between metropolitan and nonmetropolitan areas within the periphery will converge toward the within-core migration pattern; and (2) migration between core and periphery metropolitan areas will be greater in the later period.

The above discussion focuses only on gross migration flows within and between regions and metropolitan and nonmetropolitan areas. Now we add another dimension, namely, whether areas differing in socioeconomic development will send or receive individuals who differ in
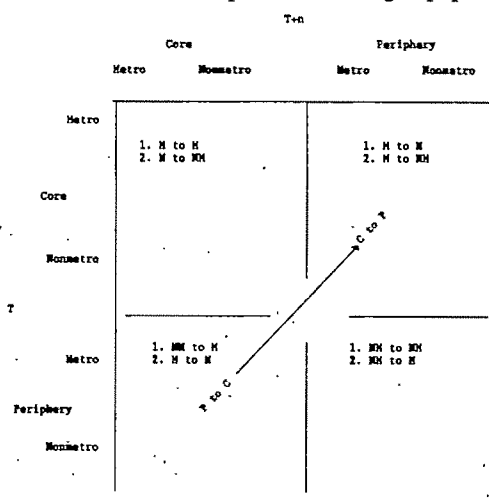
Fig. 1. Hypothesized Dominant Direction of Inter-regional and Metropolitan/Nonmetropolitan Migration Streams

attributes related to labor-force opportunities.[2] Polarized development leads to area differences in the production and utilization of human capital. Migration plays an important role here, primarily as a redistributive mechanism in which an area's stock of human capital changes in response to changing labor requirements. In addition, migration can either increase or decrease differences in the populations of sending and receiving areas, depending on the volume and direction of the flow between the areas. During the development of the core, the periphery is likely to contribute its most talented and skilled workers, reducing the pool of labor to initiate and/or sustain its own development. This drain can be reduced and subsequently reversed if the periphery is able to promote growth in new production techniques and/or goods and services (see Morgan 1975). In any event, regional differences in development and growth affect migration flows, which, in turn, transform the demographic and socioeconomic structure of the population of sending and receiving areas (see Brown and Neuberger 1977).

Differences between migrants and nonmigrants reflect area differences in socioeconomic and opportunity structures (Lee 1966). Migrant/nonmigrant education, occupation, and earnings attainment differences in rural/urban and North/South comparisons have attracted considerable attention, partly because they reveal migration patterns to areas where opportunities for socioeconomic advancement are greater (e.g., Bouvier, Macisco, and Zarate 1976; Wilson 1981, 1985). Bouvier et al. (1976) have reviewed the literature that relates migrant/nonmigrant differences in education to stages of societal development. They suggest that (1) the relative education of a population varies by societal type and socioeconomic development at origin and destination; (2) the less advanced a society, the greater the likelihood that migrants will have less education than nonmigrants at origin and destination; and (3) the more advanced the society, the greater the likelihood that migrants will have more education than nonmigrants. Figure 2 presents hypothesized migrant/nonmigrant differentials in education based on the Bouvier et al. model, according to whether the origin/destination of migrants is to or from core or periphery regions, metropolitan and nonmetropolitan areas.[3] These educational differences

Fig. 2. Hypothesized Migrant/Nonmigrant Differentials in Educational Attainment

Note: Cell values denote nonmigrant/migrant educational attainment at origin and destination. For example, (L L) indicates that the educational attainment of migrants is lower than nonmigrants at both origin and destination; while (L H) indicates that the educational attainment of migrants is lower than nonmigrants at origin, but higher than nonmigrants at destination

provide another test of the hypothesized relationship between migration and core/periphery differences in socioeconomic development discussed earlier.

The pattern in Figure 2 is based on the hypothesis that migrant/nonmigrant differences in education are a function of regional and local area differences in opportunities for schooling and employment. Specifically, the education of migrants relative to nonmigrants will be (1) high at origin and destination when both are core metropolitan and nonmetropolitan, or periphery metropolitan; (2) low at origin but high at destination when the migrant originated from the core or is involved in movement from periphery metropolitan; (3) high at origin but low at destination when the migrant is nonmetropolitan and the nonmigrant is core or periphery metropolitan, or the migrant is core nonmetropolitan and the nonmigrant is core metropolitan; and (4) low at both origin and destination when both are periphery nonmetropolitan.

The first pattern usually occurs when socioeconomic convergence reduces the impact of push

---

[2] There are likely to be important compositional differences between metropolitan/nonmetropolitan migration streams, differences that can help to illuminate the factors that give rise to and cause them to persist over time.

[3] This scheme is adapted from Bouvier et al. (1976), where metropolitan/nonmetropolitan have been substi-

tuted for rural/urban and core/periphery substituted for stages of societal development. The Bouvier et al. scheme has been modified to include several streams that they ignored, largely because these were considered rare occurrences. They suggest, for example, that, in an advanced industrial society, migration between rural areas and from urban to rural are highly unlikely to occur, although there is an abundance of evidence to the contrary (see Fuguitt 1985; Wilson 1987). Indeed, as Zelinsky (1983) suggests, rural to urban migration will probably continue as long as the former produces a surplus of births over deaths, even after a society attains a high level of urbanization.

and pull forces on migration to a minimum, and the streams themselves become dominated by the highly skilled, professionals, and individuals with above-average income.[4] Goldscheider (1983, p. 61) suggests that, in highly industrialized societies where movement is between areas that are relatively similar in socioeconomic conditions, migration has become a screening mechanism for selective social mobility (see also Blau and Duncan 1967; Miller 1977; Featherman and Hauser 1978). The second pattern reflects movement from areas of high to low socioeconomic development. The third is the classic case of highly educated rural-born natives migrating to cities seeking better opportunities. Pattern four occurs only in migration between nonmetropolitan areas where the motivation for moving is generally noneconomic.

CURRENT STUDY

This paper explores the association of migration with metropolitanization, in the context of regional socioeconomic development from 1935 to 1980 in the U.S. I hypothesize that both migration patterns within and between the metropolitan and nonmetropolitan sectors and regions, as well as migrant/nonmigrant differentials in education, reflect historical differences in socioeconomic development and settlement patterns. The 1935–80 period is appropriate to test this hypothesis because, during this period, the hegemony of the traditional manufacturing belt (identified as the core region, or industrial heartland) began to erode seriously.

Clearly, the focus on gross flows and educational differences between migrants and nonmigrants omits issues that could be fruitfully pursued under the topic of "migration and development" (see Brown and Neuberger 1977; De Jong and Gardner 1981; Bilsborrow, McDevitt, Kossoudji, and Fuller 1987). For example, an important factor not explicitly considered

is the role that different types of migrants play in population redistribution. Thus, I ignore first-time versus repeat or chronic movers (see Miller 1977, DaVanzo and Morrison 1981, Wilson 1985).

Data and Methods

The primary sources of data are the 1/100 Public Use Microdata files (PUMS) from the 1940 and 1980 decennial censuses. Supplemental information is obtained from selected Bureau of the Census publications. The PUMS file provides a complete breakdown of the U.S. population on the basis of metropolitan/nonmetropolitan residence, which is used to classify the native-born population 25 years of age and over according to their migration status during a five-year interval. The delineation of the metropolitan universe for 1940 is derived from the SMA geography established for the 1950 census. The peculiarities and limitations of these files are discussed elsewhere (see Wilson 1986, 1987).[5] Migrants are individuals who left their state or SMSA and are classified according to whether they are metropolitan to metropolitan, metropolitan to nonmetropolitan, nonmetropolitan to metropolitan, or nonmetropolitan to nonmetropolitan movers. Nonmigrants are defined as those who did not leave their state or SMSA of residence. The SMSA residence restriction applies only in instances in which an SMSA crosses state boundaries. Excluded from the analysis are those who did not report their previous place of residence and persons who were abroad (including U.S. possessions) during the five-year interval prior to the census.

In the following analysis, the nine major census divisions are grouped into core and periphery regions. The core includes the New England, Middle Atlantic, East North Central, and Pacific divisions; the periphery includes the

---

[4] There are several corollary observations that need to be stated. First, the variety of linkages between geographic areas within a politically and economically integrated society leads over time to convergence in social and economic institutions, and thus reduces the impact of area-specific push and pull forces on the volume and composition of migration streams. Second, convergence is an equilibrium-seeking process, but one that is unlikely to achieve stationality, because external forces to which a society reacts create tensions that often necessitate modifications in its pattern of adaption. Finally, the tendency toward convergence in socioeconomic structure does not mean that all regions experience an identical set of transitions as development proceeds. This is because technological and organizational changes are ever-present, altering the conditions and context within which development occurs, the character of metropolitanization, and the rate at which it occurs.

[5] The PUMS files can only provide counts of migrants during a five-year interval. Since the current analysis is concerned principally with moves that are related to employment or career, the absence of information on the PUMS files indicating the underlying reasons for moving is a serious problem. Moreover, several of the geographic criteria used to define migration are intended to partially deal with this issue. In focusing only on moves between states and between separate metropolitan and nonmetropolitan areas, we have eliminated a substantial number of moves that are associated with family life cycle changes and the demand for housing and neighborhood services. Previous analyses indicate that approximately two-thirds of interstate moves or moves between metropolitan and nonmetropolitan areas are related to labor force activity (49 percent), school attendance (8 percent), and the armed forces (4 percent) (see Long and Hansen 1979; Wilson 1979; Long and De Are 1980; Greenwood 1981). Among persons less than 25 years of age, school attendance and the armed forces account for an even larger share of moves.

Table 1. Percentage of Division Population Concentrated in Urban Areas

Division	1920	1930	1940	1950	1960	1970	1980
Core							
New England	75.9	77.3	76.1	76.1	76.4	76.6	75.1
Middle Atlantic	75.4	77.7	76.8	80.5	81.4	81.8	80.6
East North Central	60.8	66.4	65.5	69.7	73.0	74.8	73.3
Pacific	60.5	66.6	64.9	74.4	81.1	86.2	86.6
Periphery							
West North Central	37.7	41.8	43.3	52.2	58.8	63.7	63.9
South Atlantic	31.0	36.1	38.8	49.1	57.2	64.1	67.1
East South Central	22.4	28.1	29.4	39.1	48.4	54.7	55.7
West South Central	29.0	36.4	39.8	55.6	67.7	72.7	73.4
Mountain	36.5	39.4	42.7	54.9	67.1	73.1	76.4

Source: U.S. Bureau of the Census, 1983, Table 13.

West North Central, South Atlantic, East South Central, West South Central, and Mountain divisions. The core/periphery dichotomy differs from groupings in previous works; for example, the Snow Belt/Sun Belt distinction.[6] The rationale for the current classification is that the core and periphery categories differ in historical patterns of socioeconomic development, which provides the context for interpreting migration since the Great Depression. The 1940 period provides a useful benchmark for evaluating changes in the two regions. After 1940, population and economic growth shifted to the periphery, signalling a movement toward convergence in social and economic structures.

Tables 1 and 2 show that the core achieved urban and metropolitan majorities prior to the 1920s, while the periphery achieved urban majorities in the 1940s and metropolitan majorities in the 1950s. These regional differences should have important implications for migration patterns during the period of this analysis. Moreover, in 1940, the industrial structures of the two regions differed markedly (see Table 3).[7] Agriculture constituted only 9 percent of the labor force in the core region, but 32 percent in the periphery. Extractive and transformative industries account for 37 and 21 percent respectively of the labor force in core and periphery regions. Since 1940, there has been a clear trend toward regional convergence in practically all of the major industry groups. Although the share of the U.S. labor force in the core has remained virtually constant, its share of manufacturing employment declined from 69 percent to 62 percent. Most of the change has

been internal to each region: in the core, from agriculture and extractive/transformative activities to producer services; in the periphery, from agriculture to extractive/transformative and producer services. Finally, Table 4 indicates that internal regional divisions are similar in average per capita income. It is higher in the core than in the U.S., while lower in the periphery. However, regional differences in per capita income have narrowed considerably since 1929, from 46.1 to 12.1 in 1981.

Results

The first issue to be addressed is whether the volume and direction of migration between 1935–40 and 1975–80 changed in a manner consistent with the mobility-transition hypothesis. Table 5 presents the distributions of migrants by origin/destination regions and metropolitan and nonmetropolitan areas. The distribution in the origin and destination categories for 1935–40 is consistent with the mobility hypothesis. In the core, the majority of moves originating within are to and from metropolitan areas. This is to be expected because over two-thirds of the population was concentrated in metropolitan areas. The intermetropolitan stream is dominant, followed by the metropolitan to nonmetropolitan, nonmetropolitan to metropolitan, and nonmetropolitan to nonmetropolitan stream. It is evident even during this period that metropolitan areas within the core were sending more migrants to nonmetropolitan areas than they received. This imbalance is in all likelihood a result of the peripheral expansion of metropolitan settlement patterns into nonmetropolitan territories (see Wilson 1986).

The pattern of migration originating from the periphery for 1935–40 differs in several respects from the pattern observed for the core. The percentage of persons migrating was almost two points higher, and 202,000 more persons migrated to the core than left. Hence, even during this period of reduced economic growth, the direction of movement still favored the more

[6] Several authors have noted that the Sun Belt/Snow Belt dichotomy is conceptually and methodologically flawed (see Browning and Gesler 1979; Gober 1984; Keinath 1985; Gertler 1986).

[7] The results of a log linear analysis of a classification consisting of the nine census divisions, five industry groupings, and four census years also confirmed the usefulness of the core/periphery dichotomy. The analysis is available on request.

Table 2. Percentage of Division Population Concentrated in Metropolitan Areas

Division	1920	1930	1940	1950	1960	1970	1080
Core							
New England	·74.6	75.9	75.6	77.1	79.6	79.1	76.6
Middle Atlantic	79.1	82.2	82.5	83.0	83.5	89.2	87.7
East North Central	51.9	59.4	59.3	63.6	67.1	79.4	78.1
Pacific	59.7	68.7	66.2	73.9	80.1	90.3	89.4
Periphery							
West North Central	26.5	30.7	32.6	37.1	43.3	53.1	53.3
South Atlantic	21.5	28.9	31.2	38.6	48.7	70.2	˙70.4
East North Central	17.5	21.6	23.4	29.6	36.0	52.9	51.9 ·
West North Central	19.7	26.5	29.6	39.7	53.5	68.5	70.4
Mountain	14.7	17.4	20.0	30.3	48.8	61.2	63.1

Source: 1920–1960, Robinson (1968), Tables 2 and 11; 1970–80, U.S. Bureau of the Census (1981), Table 5.

economically advanced region. In addition, one can note that migration to the core was metropolitan-dominant, while migration within the periphery was nonmetropolitan-dominant. The inter-nonmetropolitan stream was at least twice the size of the three other streams, which indicates the limited availability of employment opportunities in metropolitan areas. That a slightly higher percentage of periphery migrants moved to metropolitan areas in the core region than within the periphery is also consistent with this observation.

The pattern of migration within and between the two regions during 1975–80 was markedly different than in 1935–40. The percentage migrating increased to roughly 9 percent for both regions. Moreover, 55 percent of all interstate moves originating from the core were to periphery destinations, an increase of 20 percent over the 1935–40 level. On the other hand, only 32 percent of movers from the periphery went to the core, a decline of 7

percent. Both patterns reveal that the growth momentum had shifted to the periphery region. That economic growth in the periphery was probably concentrated in metropolitan areas is indicated by the overwhelming dominance of the intermetropolitan stream regardless of whether it originated internally or from the core region. The overall pattern of migration within the periphery is becoming similar to that in the core region. This is indicated by the size ordering of the within-periphery migration streams—inter-metropolitan, metropolitan to nonmetropolitan, nonmetropolitan to metropolitan and inter-nonmetropolitan. This ordering reflects the emerging structure of its metropolitan system and its converging toward the pattern of the core.

The above results on gross migration flows within and between metropolitan and nonmetropolitan areas for core and periphery support the hypothesis of a relationship between migration, metropolitanization, and socioeconomic

Table 3. Distribution of Employment By Major Industry Groupings for Core and Periphery Regions: 1940–1970

Region/Industry[a]	Percentage of Employment							
	Within Region				Total			
	1940	1950	1960	1970	1940	1950	1960	1970
Core								
TOTAL	100.00	100.00	100.00	100.00	57.16	57.97	58.51	57.88
Agriculture	9.33	6.40	3.90	2.37	27.75	29.65	33.47	37.32
Extractive/Transformative	36.85	38.95	38.75	33.95	69.28	67.42	66.27	61.91
Distributive	10.41	11.78	10.54	10.51	61.28	62.44	59.23	58.06
Producer	14.24	15.15	19.41	25.85	63.65	61.82	60.96	59.95
Others	29.17	27.72	27.40	27.32	59.81	58.89	53.54	54.25
Periphery								
TOTAL	100.00	100.00	100.00	100.00	42.84	42.03	41.49	42.11
Agriculture	32.41	20.95	10.93	5.46	72.25	70.35	66.53	62.68
Extractive/Transformative	21.80	25.95	27.79	28.71	30.72	32.58	33.73	38.09
Distributive	8.79	9.78	10.22	10.43	38.72	37.56	40.77	41.94
Producer	10.85	12.91	17.53	23.73	36.35	38.18	39.04	40.05
Others	26.96	30.41	33.53	31.67	40.19	41.11	46.46	45.75

Source: Bureau of Economic Analysis, U.S. Department of Commerce, 1975, Regional Employment by Industry, 1940–1970.

[a] Industry Groupings, adapted from J. Singlemann (1979), consist of the following two-digit SIC Codes: Agriculture, 01–09; Extractive and Transformative, 10–39; Distributive, 40–49, 50; Producer, 60–67, 73–76, 80–86, 89; Others, 54–59, 70–72, 78–79, 91–93, and Industries not reported.

Table 4. Average Divisional Per Capita Income as a Percentage of the U.S. Average: 1929–1981

Region	1929	1944	1954	1964	1976	1981
Core						
Average	118	114	109	106	103	103
New England	112	101	99	98	97	98
Middle Atlantic	136	121	116	114	108	108
East North Central	109	107	108	105	104	100
Pacific	115	127	112	108	105	106
Periphery						
Average	71	81	83	85	90	91
West North Central	76	87	90	88	92	97
South Atlantic	86	88	90	92	98	96
East South Central	49	63	63	73	80	77
West South Central	59	74	76	79	87	91
Mountain	86	94	97	92	94	95

Source: Weinstein, Gross, and Rees, 1985, Table 2.2

[a]Averages are computed over divisions.

development. Now, I extend the evaluation of this hypothesis to include migrant/nonmigrant educational differentials. Differences in socioeconomic development between the regions should be reflected in migrant/nonmigrant educational differences, as hypothesized in Figure 2. Specifically, individuals migrating between areas of similar development should have educational levels similar to those of nonmigrants, while migrants between areas that differ in socioeconomic development should differ in education from nonmigrants. Furthermore, dif-

Table 5. Metropolitan and Nonmetropolitan Migration Flows Within and Between Core and Periphery Regions for the Population 25 Years of Age and Over at the Beginning of a Migration Interval: 1975–80 and 1935–40

Migration Streams by Region	1975–80 Number (1,000's) (1)	Percent (2)	Percent (3)	1935–40 Number (1,000's) (4)	Percent (5)	Percent (6)	Change (7) = (3) − (6)
CORE							
Total	54,549	100.00		28,447	100.00		
Nonmovers	50,472	92.53		27,480	96.60		
Movers	4,077	7.47	100.00	967	3.40	100.00	
To Core	1,853		45.45	637		65.87	−20.42
Metro to Metro	1,296		31.79	355		36.71	−4.92
Metro to Nonmetro	288		7.06	121		12.51	−5.45
Nonmetro to Metro	187		4.59	91		9.41	−4.82
Nonmetro to Nonmetro	82		2.01	70		7.24	−5.23
To Periphery	2,224		54.54	330		34.13	20.41
Metro to Metro	1,412		34.63	153		15.82	18.81
Metro to Nonmetro	484		11.87	103		10.65	1.22
Nonmetro to Metro	215		5.27	33		3.41	1.86
Nonmetro to Nonmetro	113		2.77	41		4.24	1.47
PERIPHERY							
Total	46,919	100.00		25,880	100.00		
Nonmovers	43,239	92.16		24,514	94.72		
Movers	3,680	7.84	100.00	1,366	5.28	100.00	
To Core	1,190		32.34	532		38.95	−6.61
Metro to Metro	731		19.86	181		13.25	6.61
Metro to Nonmetro	143		3.89	50		3.66	0.23
Nonmetro to Metro	241		6.55	169		12.37	−5.82
Nonmetro to Nonmetro	75		2.04	132		9.66	−7.62
To Periphery	2,490		67.66	834		61.05	6.61
Metro to Metro	1,280		34.78	164		12.01	22.77
Metro to Nonmetro	479		13.02	129		9.44	3.58
Nonmetro to Metro	410		11.14	166		12.15	−1.01
Nonmetro to Nonmetro	321		8.72	375		27.45	−18.73

Table 6. Age-Adjusted Percentage Distribution of the Population 25 Years of Age and Over by Years of Schooling Completed for Areas: 1975–80 and 1935–40

Years of Schooling Completed	1975–80				1935–40			
	Core		Periphery		Core		Periphery	
	Metro	Nonmetro	Metro	Nonmetro	Metro	Nonmetro	Metro	Nonmetro
<9 Years	13.52	18.04	19.05	29.25	58.80	65.02	61.44	73.95
9–11 Years	15.72	16.10	16.28	17.73	14.66	14.58	12.59	11.22
High School	38.88	42.53	33.21	31.50	15.39	11.67	13.98	7.39
1–3 Years College	17.34	13.52	17.13	12.61	6.00	5.70	7.10	5.00
4+ Years College	14.57	9.80	14.34	8.65	5.16	3.03	4.89	2.44
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

ferences should be greater for interregional migrants, but decline over time.

The previous discussion omitted two important considerations. First, migrants tend to be much younger and, therefore, more highly educated than nonmigrants (see Long and Hanson 1977; Wilson 1981, 1985; Sandefur 1985). In addition, age captures a "vintage" effect because each successive birth cohort receives more education than preceding cohorts. If subsequent cohorts complete more years of schooling and are also larger in size, as is true of the 1946–1961 birth cohorts, then the migration rate will increase, *ceteris paribus* (see Long and Hansen 1977; Wilson 1983).

Second, implicit in the hypothesis regarding migrant/nonmigrant educational differences is the assumption that regions have markedly different opportunity structures and labor requirements.[8] The data in Table 6, which are age-adjusted years of schooling completed by the population 25 years of age in 1935 and 1975, tend to support this observation.[9] While a dramatic upward shift in years of schooling completed occurred in all areas since 1935, the percentage in the higher-schooling categories is greater for metropolitan areas and the core region.

The ordering of the areas for both years is core metropolitan, periphery metropolitan, core nonmetropolitan, and periphery nonmetropolitan. These area educational differences indicate

that the greater the education of origin population, the greater the education of migrants from that origin. Hence, migrant/nonmigrant differences in education not only reflect the greater propensity of the highly educated to migrate, but also the education of the population of origin and destination.

Goodman's (1972) log linear model for the analysis of cross-classification is used to analyze migrant/nonmigrant differences in education net of age. The estimated model compares the education of migrants and nonmigrants separately for the 1935–40 and 1975–80 migration intervals:

$$\log (E) = \tau E + \tau EA + \tau EG + \tau EAG \quad (1)$$

where E is the log odds of completing one or more years of college versus high school or less); A is age at the beginning of a migration interval (25–34; 35–44; 45–54; 55+); and G is a sixty-four-category nested classification consisting of migration status (migrant versus nonmigrant), region (core versus periphery), metropolitan status (metropolitan versus nonmetropolitan), migration stream (metro to metro, metro to nonmetro, nonmetro to metro, and nonmetro to nonmetro), and origin/destination areas. The population less than 25 years of age at the beginning of a migration interval is excluded to eliminate the migration of dependent children, individuals who moved to attend school, and many of the young entering or leaving the armed forces.

The parameter specifications in equation (1) allow a direct test of the hypothesized migrant/nonmigrant education differentials presented in Figure 2. The additive log odds coefficients for the nested classification, G, form the basis of our discussion. These coefficients, which are reported in the appendix, give the log odds of a migrant and nonmigrant having completed one or more years of college by area of origin and destination, separately for the 1935–40 and 1975–80 migration intervals. The nested classification offers the advantage of comparing migrants and nonmigrants under all possible com-

---

[8] In this regard, the occupational distribution of a place would be a more suitable measure of opportunity structure; unfortunately, the PUMS file provides information only on an individual's usual occupation at the time of the census. Hence, the temporal ordering of migration and being employed in a particular occupation becomes difficult, if not impossible, to disentangle because of the absence of information on labor force activity prior to the move. In addition, the substantial changes made in the occupational codes reported for the 1980 census make trend analysis with previous censuses very difficult, except for very broad occupational groupings.

[9] The percentage distributions reported in Table 6 have been adjusted for the effects of age, using a procedure suggested by Kaufman & Schervish (1986).

binations of origin and destination areas. Thus, it is possible to compare, for example, the education of migrants with that of nonmigrants at origin and destination, taking into account differences in the age distribution of the two areas.

Table 7 reports migrant/nonmigrant differences in college attainment by origin, destination, and time period. These coefficients were obtained by taking the difference between the raw log odds coefficients for migrants and nonmigrants reported in the appendix, and can be interpreted as the log odds of a migrant (relative to a nonmigrant) completing one or more years of college. Positive values indicate that the college attainment of migrants is greater than that of nonmigrants. All of the coefficients except one are at least four times the size of their standard error.

The coefficients in Table 7 indicate that migrants have a greater likelihood of completing one or more years of college; this finding is consistent with previous studies, indicating a positive association between migration and education (see Long and Hansen 1977; Wilson 1981, 1985; Sandefur 1985).[10] Moreover, since migrants have completed more years of schooling than nonmigrants, regardless of area of origin or destination, the hypothesized patterns in Figure 2 have to be modified to indicate that *relative* rather than *absolute* differences are more important. The difference values in columns (3) and (6) of Table 7 indicate that migrant/nonmigrant differences vary substantially between areas of origin and destination. The zero values are of no concern, since they simply indicate, in the case of within-area contrasts, that migrant and nonmigrants at origins are the same persons as those being compared at destination. The size and signs of the other coefficients do indeed form an interpretable pattern, which mirrors the pattern of differences in the education distribution of areas in Table 6. The coefficients in Table 7 indicate that migrant/nonmigrant differences in education are (1) similar at origin and destination when both are at similar levels of socioeconomic development; (2) lower at the origin if the origin is higher in socioeconomic development; and (3) higher at the origin if the destination is higher in socioeconomic development.

---

[10] The positive association between migration and education attainment is explainable as originating from three sources. First, educational attainment increases an individual's ability to both obtain and analyze information from a multiplicity of sources, on a variety of locations, which would be of great benefit in assessing whether a move will lead to advancement. Second, education inculcates value and skills that ease the severance and establishment of social ties, and pursuing it often requires leaving the parental place of residence. Finally, the labor market area for persons with differing levels of skills tends to vary from the local to regional to national level.

A few examples may highlight the pattern of results in Table 7. First, the migrant/nonmigrant education differences are lower at origin than destination for all the metropolitan to nonmetropolitan migration streams, while the opposite is true for nonmetropolitan to metropolitan migration streams. The size of the differences, though small, does vary, depending on whether the origin or destination of the stream is within or between regions. Second, migrant/nonmigrant educational differences are slightly larger at destination than origin if the latter is core metropolitan. Finally, in most of the contrasts, while substantial declines occurred in the sizes of migrant/nonmigrant differences in education [columns (1) and (2) versus (4) and (5)], origin and destination differences changed only slightly. This pattern is clear among the four core-to-periphery streams, which indicates a substantial decline in the educational selectivity of migration from the core. A possible interpretation of this change is that the education selectivity of a stream diminishes with time as the volume increases and as socioeconomic conditions in origin and destination areas converge (see Lee 1966; Campbell and Johnson 1976; Wardwell 1980).

DISCUSSION

This paper focused on the association of migration with regional variation in metropolitanization and socioeconomic development. I attempted to determine whether the hypothesis of the mobility transition was applicable to U.S. post-depression trends in migration within and between core and periphery regions. Gross migration flows during 1935–40 and 1975–80 support the mobility transition hypothesis that developmental impulses, operating primarily through metropolitanization, structure the context of migration. Transformations in the socioeconomic structure of the periphery since the Great Depression reveal a convergence, eroding the differences in migration patterns between the periphery and core. It would be a mistake, however, to assume that convergence will eventually lead to a stationary state; the socioeconomic structure of the core continues to change and may undergo an economic transformation that may set a course different from that based on its past development (see Harrison 1982; Frey 1987; Noyelle and Stanback 1985; Weinstein, Gross, and Rees 1985; Gertler 1986; Scott 1986). Clearly, in societies with highly integrated political and economic structures, changes in core or periphery regions likely are mutually reinforcing. Although there may be similarities in socioeconomic structures and population movement, the development of the U.S. periphery is not following the trajectory of the core largely because development does not proceed as it did a century ago (see footnote 4).

Table 7. Partial Log Odds of a Migrant (Versus a Nonmigrant) having Completed One or More Years of College by Origin/Destination and Stream Type: 1975–80 and 1935–40

	1975–1980			1935–1940		
Stream Type	Origin (1)	Destination (2)	Difference (3)	Origin (4)	Destination (5)	Difference (6)
CORE						
To Core						
Metro to Metro	.2569	.2569	0.0000	.4327	.4327	0.0000
Metro to Nonmetro	.1416	.3659	−.2243	.2361	.3795	−.1434
Nonmetro to Metro	.3858	.1615	.2243	.4397	.2963	.1434
Nonmetro to Nonmetro	.2425	.2425	0.0000	.1783	.1783	0.0000
To Periphery						
Metro to Metro	.1951	.2166	−.0215	.5209	.4882	.0327
Metro to Nonmetro	.0050[a]	.2987	−.2937	.4595	.6979	−.2384
Nonmetro to Metro	.2581	.0553	.2028	.4705	.2944	.1761
Nonmetro to Nonmetro	.1432	.2126	−.0694	.5484	.6434	−.0950
PERIPHERY						
To Core						
Metro to Metro	.2179	.1964	.0215	.3234	.3561	−.0327
Metro to Nonmetro	.0954	.2982	−.2028	.1161	.2922	−.1761
Nonmetro to Metro	.1944	−.0993	.2937	.3230	.0846	.2385
Nonmetro to Nonmetro	.1054	.0360	.0694	.1413	.0463	.0950
To Periphery						
Metro to Metro	.2571	.2571	0.0000	.3729	.3729	0.0000
Metro to Nonmetro	.1043	.3765	−.2722	.2640	.5351	−.2711
Nonmetro to Metro	.3358	.0636	.2722	.4344	.1633	.2711
Nonmetro to Nonmetro	.1516	.1516	0.0000	.2070	.2070	0.0000

[a] Not twice the size of its standard error.

Some writers have noted that current regional differentials in population and economic growth, resulting from uneven development, are unlikely to lead to a stable homogeneous structure, but rather (eventually) to new labor and capital mobility (see Gober 1984; Keinath 1985; Gertler 1986; Scott 1986). Yet, there appear to be general similarities in the structure and dynamic of the metropolitanization process.

Research also reveals that differences in the socioeconomic structure of the regions reflect migrant/nonmigrant differences in education. While migrants have more education than nonmigrants, their educational differences are (1) similar at origin and destination when both are at similar levels of socioeconomic development;

(2) lower at the origin if the origin is higher in socioeconomic development; and (3) higher at the origin if the destination is higher in socioeconomic development. However, the fact that area differences in education have changed little since 1940 does not mean that the lack of differences will persist into the distant future. Indeed, there is reason to believe that core/periphery differences in education were probably even greater at the beginning of this century, with most of the decline occurring before 1940 (see Featherman and Hauser 1978). There can be little doubt that the increased availability of postsecondary education in all regions and the continuous interchange between regions due to migration will eventually eliminate the remaining differences.

Appendix. Multiple Partial Additive Log Odds of Migrants and Nonmigrants Having Completed One or More Years of College by Residence Five Years Ago and Current Residence: 1935–40 and 1975–80

		Current Residence							
		Core				Periphery			
Residence 5 Years Ago		Metro Migrant	Non-migrant	Nonmetro Migrant	Non-migrant	Metro Migrant	Non-migrant	Nonmetro Migrant	Non-migrant
CORE									
Metro									
1935–40	Origin	.3491	−.0836	.1525	−.0836	.4373	−.0836	.3759	.0836
	Destination	.3491	−.0836	.1525	−.2270	.4373	−.0509	.3759	−.3220
1975–80	Origin	.3071	.0381	.1797	.0381	.2332	.0381	.0431	.0381
	Destination	.3071	.0381	.1797	−.1862	.2332	.0166	.0431	−.2556
Nonmetro									
1935–40	Origin	.2127	−.2270	−.0487	−.2270	.2435	−.2270	.3214	−.2270
	Destination	.2127	−.0836	−.0487	−.2270	.2435	−.0509	.3214	−.3220
1975–80	Origin	.1996	−.1862	.0563	−.1862	.0719	−.1862	−.0430	−.1862
	Destination	.1996	.0381	.0563	−.1862	.0719	.0166	−.0430	−.2556
PERIPHERY									
Metro									
1935–40	Origin	.2725	−.0509	.0652	−.0509	.3220	−.0509	.2131	−.0509
	Destination	.2725	−.0836	.0652	−.2270	.3220	−.0509	.2131	−.3220
1975–80	Origin	.2345	.0166	.1120	.0166	.2737	.0166	.1209	.0166
	Destination	.2345	.0381	.1120	−.1862	.2737	.0166	.1209	−.2556
Nonmetro									
1935–40	Origin	.0010*	−.3220	−.1807	−.3220	.1124	−.3220	−.1150	−.3220
	Destination	.0010*	−.0836	−.1807	−.2270	.1124	−.0509	−.1150	−.3220
1975–80	Origin	−.0612	−.2556	−.1502	−.2556	.0802	−.2556	−.1040	−.2556
	Destination	−.0612	.0381	−.1502	−.1862	.0802	.0166	−.1040	−.2556

* Coefficient is not twice the size of its standard error.

REFERENCES

Abbott, Carl. 1981. *The New Urban America*. Chapel Hill: University of North Carolina Press.

Bilsborrow, R.E., T.M. McDevitt, S. Kossoudji, and R. Fuller. 1987. "The Impact of Origin Community Characteristics on Rural-Urban Out-Migration in a Developing County." *Demography* 24:191–210.

Blau, Peter M. and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Wiley.

Bouvier, Leon F., John J. Macisco, Jr., and Alvan Zarate. 1976. "Toward a Framework for the Analysis of Differential Migration: The Case of Education." Pp. 24–36 in *Internal Migration*, edited by Anthony H. Richard and Daniel Kubat, Beverly Hills, CA: Sage Publications.

Brown, A.A. and E. Neuberger, eds. 1977. *Internal Migration: A Comparative Perspective*. New York: Academic Press.

Brown, L.A. and R.L. Sanders. 1981. "Toward a Development Paradigm of Migration with Particular Reference to Third World Settings." Pp. 149–185 in *Migration Decision Making*, edited by G.F. De Jong R.W. Gardner. New York: Pergamon Press.

Browning, C.E. and W. Gesler. 1979. "The Sunbelt-Snowbelt: The Case of Sloppy Regionalizing." *Professional Geographer* 31:66–74.

Campbell, R. and D.M. Johnson. 1976. "Propositions on Counterstream Migration." *Rural Sociology* 41:127–45.

DaVanzo, Julie S. and Peter A. Morrison. 1981. "Return and Other Sequences of Migration in the United States." *Demography* 18:85–101.

De Jong, G.F. and R.W. Gardner, eds. 1981. *Migration Decision Making*. New York: Pergamon Press.

Featherman, David L. and Robert M. Hauser. 1978. *Opportunity and Change*. New York: Academic Press.

Frey, W.H. 1987. "Migration and Depopulation of the Metropolis: Regional Restructuring or Rural Renaissance?" *American Sociological Review* 52:240–57.

Friedmann, J. 1973. *Urbanization, Planning, and National Development*. Beverly Hills: Sage Publication.

Fuguitt, G.V. 1985. "The Nonmetropolitan Population Turnaround." *Annual Review of Sociology* 11:259–80.

Gertler, M.S. 1986. "Discontinuities in Regional Development." *Environment and Planning D: Society and Space* 4:71–84.

Gober, P. 1984. "Regional Convergence Versus Uneven Development: Implications for Sunbelt SMSA's." *Urban Geography* 5:130–45.

Goldscheider, Calvin. 1983. "Modernization, Migration, and Urbanization." Pp. 47–68 in *Population Movements: Their Forms and Functions in Urbanization and Development*, edited by Peter Morrison. Liège, Belgium: Ordina Editions.

Goodman, L.J. 1972. "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables." *American Sociological Review* 37:28–46.

Greenwood, Michael J. 1981. *Migration and Economic Growth in the United States: National, Regional, and Metropolitan Perspectives*. New York: Academic Press.

Harrison, B. 1982. "The Tendency Toward Instability and Inequality Underlying the 'Revival' of New England." *Papers of the Regional Science Association* 50:41–65.

Kaufman, Robert L. and Paul G. Schervish. 1986. "Using Adjusted Crosstabulations to Interpret Log-

linear Relationships." *American Sociological Review* 51:717–33.

Keinath, W.F. Jr. 1985. "The Spatial Component of the Post-Industrial Society." *Economic Geography* 61:223–40.

Lee, Everett, S. 1966. "A Theory of Migration." *Demography* 3:47–57.

Long, John F. 1981. *Population Deconcentration in the United States*. Special Demographic Analysis, CDS–81–5, U.S. Bureau of the Census, Washington, DC: U.S. Government Printing Office.

———. 1985. "Migration and the Phases of Population Redistribution." *Journal of Development Economics* 17:29–42.

Long, L.H. and Diana De Are. 1980. *Migration to Nonmetropolitan Areas*. U.S. Bureau of the Census, Special Demographic Analysis: CDS–89–2. Washington DC: U.S. Government Printing Office.

Long, L.H. and Kristin A. Hansen. 1977. "Migration Trends in the United States since 1935." Paper presented at the annual meeting of the International Union for the Scientific Study of Population. Mexico City.

———. 1979. *Reasons for Interstate Migration*. Current Population Reports: Special Studies: Series P-23, No. 81.

Miller, Ann. 1977. "Interstate Migrants in the Untied States: Some Social-Economic Differences by Type of Move." *Demography* 14:1–17.

Mollenkopf, J.H. 1983. *The Contested City*. Chapel Hill: University of North Carolina Press.

Morgan, D.T. 1975. "Growth Pole Theory, Technological Change, and Regional Economic Growth." Papers and Proceedings of the Regional Science Association 16:195–212.

Morrill, R.L. 1980. "The Spread of Change in Metropolitan and Nonmetropolitan Growth in the United States, 1940–1976." *Urban Geography* 1:118–29.

Norcliffe, Glen B. 1984. "Nonmetropolitan Industrialization and the Theory of Production." *Urban Geography* 5:25–42.

Noyelle, Thierry J. and Thomas M. Stanback, Jr. 1985. *The Economic Transformation of American Cities*. New York: Rowman and Allanheld.

Robinson, W.C. 1968. "Metropolitan and Urban Growth in the U.S., 1900–1960." Information Report Number 58, The Institute for Research on Land and Water Resources, The Pennsylvania State University, University Park, PA.

Sandefur, G.D. 1985. "Variations in Interstate Migration of Men Across the Early Stages of the Life Cycle." *Demography* 22:353–66.

U.S. Bureau of the Census. 1981. 1980 Census of Population: Vol. I, Characteristics of the Population, PC 80–1–A. Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census. 1983. 1980 Census of Population: Vol. I, Characteristics of the Population, Number of Inhabitants, PC 80–1–A1. Washington, D.C.: U.S. Government Printing Office.

Scott, A.J. 1986. "Industrialization and Urbanization: A Geographical Agenda." *Annals of the Association of American Geographers* 76:25–37.

Singlemann, J. 1979. *From Agriculture to Services*. Beverly Hills, CA: Sage Publications.

Smith, N. 1984. "Deindustrialization and Regionaliza-

tion: Class Alliance and Class Struggle." Papers of the Regional Science Association 54:113–28.

Vining, Daniel R. Jr. 1986. "Population Redistribution Towards Core Areas of Less Developed Countries, 1950–1980." *International Regional Science Review* 10:1–45.

Wardwell, J.M. 1980. "Toward a Theory of Urban-Rural Migration in the Developed World." Pp. 71–114 in *New Directions in Urban-Rural Migration: The Population Turnaround in Rural America,* edited by David L. Brown and John M. Wardwell. New York: Academic Press.

Weinstein, B.L., H.T. Gross, and J. Rees. 1985. *Regional Growth and Decline in the United States,* New York: Praeger.

Wilson, F.D. 1979. *"Residential Consumption, Economic Opportunities and Race."* New York: Academic Press.

———. 1981. "Migration and Socioeconomic Attainment." Discussion Paper No. 664–81, Institute for Research on Poverty, University of Wisconsin, Madison.

———. 1983. "Cohort Size Effects and Migration." *International Migration Review* 17:484–504.

———. 1984. "Urban Ecology: Urbanization and Systems of Cities." *Annual Review of Sociology* 10:283–307.

———. 1985. "Migration and Occupational Mobility: A Research Note." *International Migration Review* 19:278–92.

———. 1986. "Temporal and Subnational Variation in the Reversal of Migration Flows between Metropolitan and Nonmetropolitan Areas: 1935–1980." *Social Forces* 65:501–24.

———. 1987. "Metropolitan and Nonmetropolitan Migration Streams: 1935–1980." *Demography* 24:211–28.

Zelinsky, Wilbur. 1971. "The Hypothesis of the Mobility Transition." *Geographical Review* 61:219–49.

———. 1983. "The Impasse: Migration Theory: A Sketch Map for Potential Escapes." Pp. 19–46 in *Population Movements,* edited by Peter Morrison. Liège, Belgium: Ordina Editions.

COMMITMENT AND THE MODERN UNION: ASSESSING THE LINK BETWEEN PREMARITAL COHABITATION AND SUBSEQUENT MARITAL STABILITY*

NEIL G. BENNETT
Yale University

ANN KLIMAS BLANC
Institute for
Resource Development/Westinghouse

DAVID E. BLOOM
Columbia University

In recent years, the incidence of premarital cohabitation has increased dramatically in many countries of Western Europe and in the United States. Analyzing data from the 1981 Women in Sweden survey using a hazards-model approach, we focus on the relationship between premarital cohabitation and subsequent marital stability. Our results indicate that women who cohabit premaritally have almost 80 percent higher marital dissolution rates than those who do not. Women who cohabit for over three years prior to marriage have over 50 percent higher dissolution rates than women who cohabit for shorter durations. Last, cohabitors and noncohabitors whose marriages have remained intact for eight years appear to have identical dissolution rates after that time. We provide evidence that strongly suggests a weaker commitment to the institution of marriage on the part of those who cohabit premaritally.

INTRODUCTION

Nonmarital cohabitation is one of several nontraditional family forms and household structures increasingly observed in many developed countries. The rising propensity to cohabit outside of marriage has been linked to demographic trends such as increasing proportions never married, increases in the average age at marriage, rising divorce rates, and rising proportions of births outside of marriage. Many sociologists and demographers expect that the incidence of nonmarital cohabitation will continue to rise for many years (Davis 1983; Glick 1984; Macklin 1978; Norton 1983; Spanier 1986; Westoff 1978). Glick asserts, for example, that the number of cohabiting unmarried couples in the United States, which almost tripled in the 1970s, will have nearly doubled during the 1980s.

The increase in nonmarital cohabitation has been particularly marked in Scandinavia. In Sweden, for example, unmarried cohabiting

couples comprised only one percent of all couples in 1960. In 1970, the proportion was seven percent and in 1979, 15 percent (Trost 1980). In Denmark, between eight and nine percent of all unions were nonmarital in 1974; by 1978, it was 13 percent. A similar but somewhat less marked trend has been observed in most of Western Europe (Audirac 1982; Brown and Kiernan 1981; Festy 1980).

Clearly, understanding the links between nonmarital cohabitation and other steps in family formation and dissolution becomes increasingly important as the proportion participating in this nontraditional family form grows. This paper focuses on the relationship between premarital cohabitation and subsequent marital stability.

Two hypotheses have been raised with respect to this relationship (see, e.g., Cherlin 1981; Macklin 1978; Mead 1966). One hypothesis states that only the most stable of cohabiting couples marry. Cohabitation is viewed as a form of trial marriage in which unstable unions are "weeded out" before marriage. In unions that lead to marriage, couples have presumably adjusted to expected marital roles and can avoid pitfalls associated with marriage to people with whose living habits one is unfamiliar.

The matching process implied by this hypothesis may represent the latest stage in the historical evolution of Western marriage markets. Marriage has never been a random coupling process in Western societies. Information about potential spouses has always played an important role in the making of matches. But

the nature of the information deemed important, and the process by which it is gathered, has changed over time. Historically, family elders or the community played a dominant role in arranging marriages; the suitability of potential matches was evaluated largely in terms of individuals' social and economic backgrounds. Individuals were raised with the expectation that they would make adjustments after marriage that were necessary to ensure longlasting and beneficial unions. Over time, the bride and groom have played more prominent roles in the matching process—they collect and process much of the information about potential spouses themselves (e.g., through dating), and they tend to place greater weight on information relating to personal characteristics such as personality and physical appearance. As a practical matter, it appears that careful screening before marriage has displaced in part the willingness to make adjustments after marriage as the supposed key to promoting successful unions.

Although the emergence of premarital cohabitation lends itself to a variety of interpretations, it may be viewed as an extension of the notion that more information on a range of personal characteristics improves the quality of marital unions. Research on the validity of this perspective is difficult. On the surface, one might judge the secular increase in marital instability as evidence against this view. But this is a difficult link to establish since so many other factors affecting both the process of entry into marriage and marital stability have changed over time.

The second hypothesis regarding cohabitation and marital stability states that those who cohabit are a select group of people for whom relationships in general—both nonmarital and marital—are characterized by a lack of commitment and stability. Moreover, those who cohabit may attach less importance to traditional institutions, such as legal marriage, and may be more willing to dissolve unsatisfying relationships (see Carlson 1986). Thus, premarital cohabitors are expected to have higher marital-dissolution rates than the married population who did not cohabit. This hypothesis does not necessarily preclude the hypothesis that cohabitation selects out less compatible partners before marriage. Even if cohabitors are more likely to dissolve their marriages, they may have lower dissolution rates than they otherwise would have without cohabiting.

Although direct evidence to support this interpretation cannot be obtained from the available data, the results of several studies of cohabitation and marriage in other countries are suggestive. In a survey of 18- to 29-year-olds in France in 1977, Carlson (1986) reports that, compared to married couples who had not cohabited, those who had cohabited or were cohabiting at the time of the survey were twice as likely to view marriage as a response to social pressure and were half as likely to see marriage as the result of the couple's desire to add something to their union. In addition, when respondents were asked about the future of marriage, cohabitors were less likely to predict that marriage would continue to be the dominant form of living together and more likely to predict that marriage would eventually disappear.

Furthermore, the results of a series of surveys conducted in Denmark[1] during the 1970s (the Euro-barometer surveys) suggest that cohabiting individuals are less likely than married individuals to subscribe to traditional sex roles. For example, in comparison to married couples of the same age, those who were living together tended to be more accepting of a husband moving for his wife's job and more likely to think it reasonable for a man to perform household chores, such as cleaning and ironing. In Denmark, cohabitors are less likely to report that their religion is important to them.

We attempt to disentangle the relationship between cohabitation and marital stability by investigating the empirical validity of each of these hypotheses. Further analysis illuminates some of the complexities involved in that relationship. Because our analysis is based on individual-level data, we will be able to control for several individual-specific variables that might reasonably be related to both premarital cohabitation and marital stability. Although interpretation of our results is limited insofar as we are unable to control for all important variables, much can be learned about the process of entry into marriage and its implications for subsequent marital stability.

THE DATA

Few data sets are appropriate for testing the above hypotheses and few studies have used appropriate data and methodology.[2] However, a

---

[1] The Euro-barometer surveys are conducted by the Commission of the European Communities and made available through the ICPSR. Although nine countries participate in the surveys, the sample sizes are small and only in Denmark are there sufficient numbers of persons in the sample cohabiting to allow the construction of meaningful crosstabulations. The results reported here are drawn from Euro-barometer 3 and Euro-barometer 8.

[2] To our knowledge, virtually all of the studies that have been conducted on this subject are based on samples of *currently married* couples rather than samples of ever-married women (see, e.g., DeMaris and Leslie 1984). As a result, the least successful or stable marriages (i.e., those that have been dissolved) are not observed. Consequently, the results are biased by

1981 Swedish survey, entitled "Women in Sweden," has an extensive cohabitational and marital history as well as a pregnancy history and numerous background variables for each respondent. The survey, conducted by the Swedish National Central Bureau of Statistics (now Statistics Sweden), was based on a sample of 4,966 women aged 20 to 44 and resident in the country as of February 1981. Interviews were carried out with 4,300 respondents and took place primarily between March and May of 1981 (Statistics Sweden 1981).[3]

In the section of the survey dealing with marriage and cohabitation, respondents were asked to provide the dates (month and year) of all periods in their lives during which they "lived together with a man, either as married or without being formally married." For each period, the dates the couple "moved in together" married (if applicable), and "split up" (if applicable) were recorded. Marriages and periods of cohabitation lasting less than one month were not recorded. The date of dissolution referred to the date the couple ceased living together rather than the date of divorce. Our analysis focuses on the dissolution, as indicated by marital separation, of first marriages. Importantly, the population here examined is composed only of ever-married women. Once we established that a woman entered a first marriage, we classified her as a cohabitor if she cohabited with her first husband immediately prior to marriage. Never-married women who either were cohabiting at the time of the survey or had cohabited before the survey date were not included in our sample.

PRELIMINARY ANALYSIS

As shown in Table 1, almost two-thirds of the women in the sample cohabited (for at least one month) with their first husband immediately prior to marrying. Overall, nearly 18 percent had experienced the dissolution of their first marriage by the time of the survey. Among cohabitors, 18.3 percent had separated from their husbands, and among noncohabitors, 17.4 percent. This simple crosstabulation reveals only trivial differences between the dissolution rates of cohabitors and noncohabitors. A chi-square test is unable to reject the null hypothesis that premarital cohabitation and marital dissolution are independent events.

Table 1. Percentage of Women Experiencing Marital Dissolution by Premarital Cohabitation Experience

| | Status of First Marriage at Time of Survey | | |
	Intact	Dissolved	All Women
Did cohabit	81.7	18.3	65.0
premaritally	(1472)	(329)	(1801)
Did not cohabit	82.6	17.4	35.0
premaritally	(800)	(168)	(968)
All women	82.1	17.9	100.0
	(2272)	(497)	(2769)

Note: Numbers of cases are reported in parentheses.

The comparison of gross dissolution rates between cohabitors and noncohabitors fails to control for a key variable related to dissolution probabilities: length of exposure to the risk of divorce. This variable may be important because cohabitors tend to marry later than noncohabitors and because there has been a cross-cohort increase in the propensity to cohabit. A briefer period of exposure, all else equal, would depress the proportion of cohabitors with dissolved marriages relative to the corresponding proportion of noncohabitors.

We control for the differential exposure of cohabitors and noncohabitors to the risk of separation by constructing life tables for the two groups. These tables provide estimates of the probability that a woman will dissolve her first marriage at each duration, taking into account her length of exposure to risk (i.e., how long she has been married). Women who have dissolved a first marriage contribute exposure at each duration until the point of dissolution. Women who are still married at the time of the survey contribute exposure at each duration prior to the survey date. Life tables, therefore, incorporate information both about women who have separated and those who have not separated by the survey date.

The cumulative proportion of marriages dissolved by a given duration of marriage is shown in Figure 1 separately for cohabitors and noncohabitors. Once we account for differential exposure between cohabitors and noncohabitors, differences in marital dissolution occurring between the two groups become evident. Within 10 years of the date of their first marriage, 18 percent of the cohabitors in the sample had separated compared to only 10 percent of the noncohabitors in the sample. Within 20 years, the figures had risen to 34 and 24 percent, respectively.

Given that the cohabitors and noncohabitors differ with respect to marital stability, it is natural to explore whether these two groups of women differ in other ways as well. In particu-

---

inclusion of a disproportionately large number of the stable marriages. For a review, see Macklin (1978).

[3] The fact that we are studying women only should not, of course, indicate that women are solely responsible for the stability of their marriages. Either partner could initiate the dissolution of the marriage, although we only observe the characteristics and behavior of women.

Fig. 1. Observed Proportion of First Marriages Dissolved by Duration x

Table 3. Percentage Distribution of Duration of Premarital Cohabitation with First Husband

Duration (months)	Percent	
1–3	12.9	
4–6	10.8	
7–9	8.5	
10–12	8.4	
1–12		40.6
13–18	13.9	
19–24	10.2	
25–30	7.9	
31–36	5.1	
13–36		37.1
37–48	9.1	
49–60	5.1	
>60	8.1	
>36		22.3

lar, are there other factors that differentiate these two groups that could account for the differences in rates of marital dissolution, rendering the cohabitation factor per se insignificant?

Table 2 compares selected sample statistics of ever-married women who did and did not premaritally cohabit. Those who cohabited are younger than those who did not, somewhat more likely to have had a premarital conception, and more than twice as likely to have had a premarital birth, though less likely to have had a marital birth.

Women who lived with their first husband immediately prior to marriage displayed great variability in the length of cohabitation. Table 3 indicates that approximately two out of five such women spent one year or less living with their future spouse. About the same proportion cohabited premaritally for one to three years, and the remaining fifth or so lived with their partner for over three years before they married.

THE MODEL

It is clear from Table 2 that several factors may simultaneously affect marital dissolution rates. Consequently, we invoke a hazards-model approach, which may be thought of as a

Table 2. Sample Means of Selected Characteristics of Ever-married Women by Premarital Cohabitation Experience and Status of First Marriage at Time of Survey

	Women who Cohabited Premaritally		Women who Did Not Cohabit Premaritally	
	First marriage		First marriage	
	Dissolved	Intact	Dissolved	Intact
Age at survey -	35.1	33.1	38.3	37.3
Age at first marriage	22.0	24.0	21.1	22.3
Age at cohabitation	20.4	21.9	—	—
Premarital conception .	.64	.53	.57	.37
Premarital birth	.38	.31	.19	.10
One or more marital births	.67	.82	.86	.93
Cohabited premaritally				
more than once	.06	.09	–0–	.003*
Occupation of main breadwinner				
during childhood:				
Salaried employee	.26	.25	.28	.24
Skilled or unskilled worker	.48	.49	.43	.40
Farmer or self-employed	.26	.26	.29	.36
Education				
Less than secondary	.78	.65	.72	.75
Completed secondary	.09	.17	.11	.10
More than secondary	.13	.18	.17	.15

* Two women cohabited premaritally more than once, though they did not cohabit immediately prior to their first marriage.

multivariate extension of the simple life-table analysis presented above (see, e.g., Cox and Oakes 1984).

We assume that there is a hazard, or risk, of dissolution at each marital duration, d, and we allow this duration-specific risk to depend on individual characteristics.[4] In the proportional-hazards model, a set of individual characteristics represented by a vector of covariates shifts the hazard by the same proportional amount at all durations. Thus, for an individual i at duration d, with an observed set of characteristics represented by a vector of covariates, Z_i, the hazard function, $\mu_i(d)$, is given by

<hr>

[4] We have explored the possibility that the salient measure of duration is "duration since the initiation of the union," not "duration since the initiation of the marriage." For married women who did not cohabit premaritally, obviously the measures are identical. However, for those who did cohabit, this new duration measure reflects the total amount of time that a couple has been in a union, formal or otherwise. In this regard, we can test the hypothesis that there are no differences in marital dissolution probabilities between cohabitors and noncohabitors using this measure of duration. This hypothesis might be true if people "get tired of their partners" within some length of time, regardless of their marital status. For example, women who have been married ten months with no prior cohabitation would have dissolution rates similar to those who have been married only six months but with four months of premarital cohabitation.

We test this hypothesis by counting duration as that since union; however, we censor our data before the time of marriage. In our example, then, we would pretend to observe those women who cohabited before marriage only in their fifth month of union and beyond. In this way, the fact that these women cannot possibly divorce before they are married does not bias our results. It is important to bear in mind that our results are conditioned upon entering into marriage. In this model, we also include a covariate denoting whether one cohabited. Under this hypothesis, the cohabitation covariate should be irrelevant to the likelihood of marital dissolution. However, when we estimate a model specified in this way, the relationship between cohabitation and subsequent marital stability is in the same direction and virtually as strong in magnitude.

Further analysis may address whether a given length of premarital cohabitation could be translated into an "equivalent" length of marriage. Suppose, for example, that cohabitation were, for argument's sake, only half as "intense" a relationship as marriage with respect to the amount of deterioration (or improvement, for that matter) that a relationship may experience over time. Referring again to our example above, in this scheme we pretend to observe the cohabiting couples in what we call the third "pseudo-month." That is, their marriage occurs in their third marriage-duration-equivalent month, since the four months of cohabitation translate into two months of marriage duration. We specified a range of such translations and in no case was the cohabitation parameter estimate anything but similar to that obtained in models measuring duration as that since marriage.

$$\mu_i(d) = \exp[\lambda(d)]\exp[Z_i\beta]$$

where β is a vector of parameters and $\lambda(d)$ is the underlying duration pattern of risk. In this model, then, the underlying risk of dissolution for an individual i with characteristics Z_i is multiplied by a factor equal to $\exp[Z_i\beta]$.

We also examine a set of more general models to test for departures from some of the restrictive assumptions built into the proportional-hazards framework. More specifically, we allow for time-varying covariates (e.g., the occurrence of a first marital birth) as well as allow for the effects of individual characteristics to vary with duration of first marriage. This model may be written as

$$\mu_i(d) = \exp[\lambda(d)]\exp[Z_i(d)\beta(d)]$$

where $\lambda(d)$ is defined as in the proportional-hazards model, $Z_i(d)$ is the vector of covariates, some of which may be time-varying, and $\beta(d)$ represents vector parameters, some of which may give rise to nonproportional effects. The model parameters are estimated using the method of maximum likelihood (see Tuma 1979). The estimation procedure assumes that the hazard, $\mu_i(d)$, is constant within duration intervals. The intervals (in years) that we have chosen are 0–1, 2–4, 5–7, 8–11, and 12 and greater. Experimentation with alternative intervals yielded no substantive differences in our analysis.

RESULTS AND DISCUSSION

The object of this statistical analysis is to identify the direction and magnitude of the relationship between premarital cohabitation and the risk of marital dissolution controlling for other factors associated with marital disruption. The first model that we report includes all covariates available in the Swedish survey that we could sensibly hypothesize to relate to marital dissolution (see Becker, Landes, and Michael 1977; Cherlin 1977; Menken, Trussell, Stempel, and Babakol 1981; Morgan and Rindfuss 1985; Murphy 1985; Teachman 1982; Waite, Haggstrom, and Kanouse 1985). Table 4 reports the antilogs of the parameter estimates and the t-statistics associated with the parameter estimates themselves, based on a simple proportional-hazards model.[5] Because the estimates

<hr>

[5] The cross-sectional nature of the data will generate a bias by selecting women with short duration of cohabitation. In addition, we will tend not to observe women who marry late, either due to long premarital cohabitation or other factors. One approach to this issue is to note that the selection problem is less important for older women because people reach their "final state"

Table 4. Monthly Hazard of Marital Dissolution and Covariate Effects—Preliminary Model

Covariate	Parameter Antilog (t-statistic)
Premarital cohabitation = 1–3 months	1.643 (3.16)
Premarital cohabitation >3 months	1.790 (5.56)
Premarital cohabitation >1 time	1.355* (1.20)
Age at marriage <21 years	1.913 (6.80)
Premarital birth	1.518 (3.97)
First marital birth	.749 (2.37)
*Education***	
Completed secondary	1.077* (0.47)
More than secondary	1.175* (1.16)
*Occupation of main breadwinner during childhood****	
Skilled or unskilled worker	1.003* (0.02)
Salaried employee	1.460 (3.02)

Duration (in years)	Monthly Hazard (t-statistic)
0–1	.000517 (44.3)
2–4	.000811 (41.9)
5–7	.000713 (38.3)
8–11	.000785 (37.8)
≥12	.000940 (37.9)

Number of observations = 2,769
log likelihood = − 3676.701

Note: t-statistics in this and subsequent tables refer to the actual parameters and not their antilogs.

* Estimate not significantly different from zero at the .05 level.

** Omitted category is "less than secondary school graduate."

*** Omitted category is "farmer or self-employed."

eventually. This suggests a test for the selection hypothesis. One would fit the model to separate cohorts and examine whether the relationship between divorce and cohabitation differs among the various cohorts. The test is not entirely satisfactory because looking at differences across cohorts could be interpreted either as evidence of selection bias or of genuine behavioral changes. Nonetheless, performing such an analysis yielded results that are not consistent with the hypothesis suggesting cross-cohort differences.

are maximum likelihood, they are asymptotically normally distributed, facilitating the drawing of statistical inferences.

First, we categorize women according to their premarital cohabitation experience: those who did not cohabit with their first husband (or did so for less than one month); those who cohabited one to three months; and those who cohabited more than three months. No premarital cohabitation is the omitted category. This categorization is intended to test the hypothesis that women who cohabit for very short durations are more similar to those who do not cohabit at all than they are to longer-term cohabitors. Those who cohabit for a short time may be either formally or informally engaged and do so merely for logistical reasons, having at the outset already committed themselves to marrying. Instead, we find that, compared to noncohabitors, those who live together before marriage for either a brief or extended period of time are to a similar extent much more likely to dissolve their marriages (the parameter estimates for the two groups of cohabitors are not significantly different). In subsequent models, we combine all cohabitors into one group.

The overall association between premarital cohabitation and subsequent marital stability is striking. The dissolution rates of women who cohabit premaritally with their future spouse are, on average, nearly 80 percent higher than the rates of those who do not. (Recall from Table 3 that only one-eighth of women cohabiting do so for only one to three months.) This finding is comparable to that found by Blanc (1985) for Norway and by Balakrishnan, Rao, Lapierre-Adamcyk, and Krotki (1987) for Canada.[6] The magnitude of the cohabitation parameter is slightly smaller than that of age at marriage and greater than that of a premarital birth. Note that the covariate that indicates a woman cohabited more than once before marriage is positive but not significant. We may conclude, then, that the higher dissolution rates of cohabitors do not stem from a small group of "repeat cohabitors" who have especially low commitment to the institution of marriage and to relationships in general.

Age at marriage has been dichotomized into those who married at less than 21 years of age and those who married at age 21 or older. We

---

[6] Balakrishnan et al. (1987) find 50 percent higher dissolution rates for cohabitors using a proportional-hazards model in which several other variables are incorporated. Using life tables, Blanc (1985) finds that in Norway the cumulative proportion of first marriages ending in separation after five years is .12 for cohabitors and .06 for noncohabitors who married before age 21. For women who married at age 21 or later, the corresponding proportions are .06 and .02.

also include a covariate that indicates whether a woman had a birth prior to her first marriage. The event of a first marital birth is entered as a time-varying covariate (i.e., its value varies with duration) that assumes the value 0 at each duration until the first birth within marriage occurs and 1 at each duration thereafter. The coefficient may be interpreted as the relative risk of marital dissolution for a woman who has had a first birth, subsequent to that birth, compared to the corresponding risk for women who had not yet had a marital birth.

Additional results in Table 4 show that women who marry at a relatively young age or have a premarital birth have substantially higher marital dissolution rates than those who defer marriage or restrict their childbearing to within marital unions. Women who marry early appear to have almost double the rate of dissolution of their late-marrying counterparts; women with a premarital birth have a rate that is one-half higher than women who did not give birth premaritally. However, the first birth within marriage tends to have a stabilizing effect on the marriage; dissolution rates of women who give birth within marriage are one-quarter lower subsequent to the birth compared to those women at the same marriage duration who have not given birth.

These results are not surprising, as they are consistent with previous research. For example, in his analysis of marital disruption in Great Britain, Murphy (1985) found that for every year that age at marriage is reduced, the risk of dissolution increases by 16 percent. Similarly, Menken et al. (1981) found that for both white and black women in the United States, separation rates decline regularly with increasing age at marriage. The occurrence of a premarital birth has also been found to have a significant positive effect on the rate of marital disruption (Menken et al. 1981; Morgan and Rindfuss 1985; Teachman 1982). Although evidence regarding the relationship between marital fertility and marital dissolution is somewhat unclear, our findings are consistent with recent studies that suggest, at least for the first birth, that this relationship is negative (Becker et al. 1977; Teachman 1982; Thornton 1977; Waite et al. 1985).

Level of education has been found to be negatively correlated with the likelihood of divorce (Menken et al. 1981; Morgan and Rindfuss 1985; Teachman 1982). In Sweden, other factors appear to vitiate any bivariate relationship that may exist between education and marital instability.[7]

Social background, which is measured here by the occupation of the "main breadwinner" in the respondent's childhood home, may be an indicator of several factors including type of education, labor-force participation, and parent's marital status (see Bernhardt and Hoem 1985). Those women who grew up in a household in which the main breadwinner had been a salaried employee (i.e., a white-collar worker) have substantially higher marital dissolution rates than other women.[8]

Although the relationship between marital dissolution and each of our fixed covariates may change with marital duration (see Morgan and Rindfuss 1985), estimation of models in which these covariates are allowed to vary with duration shows that only premarital cohabitation has significant duration-dependent effects.

Women who cohabit prior to marriage may well be a group that is heterogeneous with respect to characteristics that were not measured in the Swedish survey (e.g., in their level of religiosity, personal maturity, or the stability of their parents' marriage). Thus, we may hypothesize in very simple terms, for example, that this group is composed of two subgroups: those who believe more and less in the institution of marriage as a lifetime commitment. Given this hypothesis, the "less-committed" group may be expected to dissolve their marriages at a relatively high rate, leaving behind the "more-committed" group (which has dissolution rates indistinguishable from the group that did not cohabit). If this is the case, then we would expect the relationship between cohabitation and dissolution to decrease in magnitude across duration.

Indeed, this more refined hypothesis is borne out by the results shown in Table 5 (variables found insignificant in the previous model are omitted). The relationships between marital stability and age at marriage, whether one had a premarital birth, and the timing of the first marital birth remain qualitatively identical to those observed in the previous model. However, the nature of the relationship between marital stability and whether a woman cohabited with her future spouse changes substantially with marital duration. The monthly hazard of marital dissolution in the first two years of marriage for

---

[7] We should note that this variable measures the respondent's level of education at the time of the survey, not at marriage. See Hoem (1986) for a detailed

discussion of the problems in the information on completed education in the Swedish survey.

[8] This finding regarding main breadwinner's occupation is somewhat puzzling. It is possible that the mothers in these households were more likely to have worked outside the home and were themselves subject to higher dissolution rates. This behavior may to some extent be transmitted across generations. Unfortunately, given the available data, we are unable to test this or related hypotheses.

those women who have cohabited premaritally is over three times that of those who have not. The hazard for cohabitors declines to approximately two times that of noncohabitors in the interval from two to eight years of marriage. After the first eight years of marriage, marriage dissolution rates of cohabitors and noncohabitors converge to the extent that any differences are small and statistically insignificant.

It is not possible to determine conclusively whether one should interpret this pattern of changing effects across duration from a life-course perspective assuming a homogeneous cohort of women or from a perspective that incorporates the notion of heterogeneity. From a life-course perspective, all couples who cohabit prior to marriage may be equally likely to dissolve their marriages at a relatively high rate during the first several years of marriage. After this time, however, couples who remain in intact marriages "settle in" and have dissolution rates essentially the same as those of couples who did not cohabit premaritally.

As outlined earlier, an alternative interpretation views those who cohabit as a group that is heterogeneous with respect to one or more unobserved characteristics that are associated with the probability of dissolution. Thus, after the first eight years of marriage, those women with a greater propensity to divorce—due to the various characteristics that we have not observed—are selected out. The subgroup of women remaining is indistinguishable from the segment of the population that never cohabited.

The duration of cohabitation is one characteristic that varies among cohabitors (see Table 3), ranging from one month to more than ten years, with a mean cohabitational spell of approximately two years. The results presented in Table 6 derive from a model in which we examine only the premarital cohabitors in our sample. We include the duration of premarital cohabitation as a covariate to evaluate two hypotheses. First, it is possible that couples who cohabit for only a short period of time before marriage, in contrast to long-term cohabitors, have less opportunity to develop an understanding of each other and to recognize and resolve potential conflicts. Should this be the case, we would expect the duration of cohabitation to be negatively related to the rate of dissolution.

Alternatively, couples who cohabit for a long period of time may be those in which one or both partners are unsure about, or ideologically opposed to, the institution of marriage itself, but who marry perhaps due to mounting external pressure. Furthermore, individuals who live together for several years before marrying may become accustomed to a relatively individualistic mode of behavior (see Rosenblatt and Budd

Table 5. Monthly Hazard of Marital Dissolution and Covariate Effects—Premarital Cohabitation as a Duration-dependent Parameter

Covariate	Parameter Antilog (t-statistic)
Age at marriage <21 years	1.879 (6.74)
Premarital birth	1.494 (3.94)
First marital birth	.744 (2.42)
Occupation of main breadwinner during childhood = salaried employee	1.493 (3.87)

Duration (in years)	Monthly Hazard (λ) (t-statistic)	Premarital Cohabitation Parameter Antilog (t-statistic)
0–1	.000339 (25.8)	3.226 (3.57)
2–4	.000818 (35.2)	1.899 (3.21)
5–7	.000623 (30.5)	2.294 (3.40)
8–11	.000934 (34.4)	1.412* (1.64)
≥12	.001107 (36.6)	1.341* (1.49)

Number of observations = 2,769
log likelihood = − 3674.085

* Estimates not significantly different from zero at the .05 level.

Table 6. Monthly Hazard of Marital Dissolution and Covariate Effects—Only Women who Cohabited Premaritally

Covariate	Parameter Antilog (t-statistic)
Age at marriage <21 years	2.079 (6.08)
Premarital birth	1.384 (2.64)
First marital birth	.714 (2.38)
Occupation of main breadwinner during childhood = salaried employee	1.534 (3.32)
Duration of Cohabitation	
6–18 months	1.148* (0.95)
19–36 months	1.026* (0.15)
>36 months	1.541 (2.24)

Duration (in years)	Monthly Hazard (λ) (t-statistic)
0–1	.000929 (36.7)
2–4	.001358 (33.7)
5–7	.001282 (30.6)
8–11	.001206 (29.5)
≥12	.001357 (28.3)

Number of observations = 1,800
log likelihood = −2634.804

* Estimate not significantly different from zero at the .05 level.

1975). Cohabitors are known, for example, to value the independence that comes with cohabitation, which is sacrificed to some extent in marriage. That is, cohabitors are often attracted to their nonmarital arrangement precisely because they view that arrangement as one associated with greater individual freedom than would be the case with marriage (see Blumstein and Schwartz 1983). Consequently, those who cohabit premaritally for an extended period of time may miss the independence that existed in their previous arrangement more than those who live together for a relatively short time. Moreover, we might expect that long-term cohabitors have been more stigmatized due to the nonconformity implicit in their unusually long spell of cohabitation. It might be easier for them to withstand the social repercussions of divorce than it is for short-term cohabitors.

Long periods of cohabitation would thus be associated with higher rates of dissolution.

The results shown in Table 6 are consistent with the latter hypothesis. Women who cohabit premaritally for more than three years have 54 percent higher marital dissolution rates than those who cohabit for shorter durations. Those who cohabit for three years or less appear to have essentially identical rates of dissolution. (The proportional factors for categories of duration 6–18 months and 19–36 months are not significantly different from one, indicating that the dissolution rates are not distinguishable from those of women in the base category, 0–5 months.)[9]

The last model that we discuss, the parameter estimates of which are presented in Table 7, refers only to women who did not live with their prospective husbands before marriage. Comparing the results in Tables 6 and 7, we see that the relationships between three factors—age at marriage, whether one had a premarital birth, and whether the main breadwinner during one's childhood was a salaried employee—and marital dissolution are similar for cohabitors and noncohabitors.

The impact of a first marital birth, however, on marital stability subsequent to that birth is insignificant for women who did not cohabit premaritally. This result stands in stark contrast to the pronounced stabilizing effect of the first marital birth among couples who did live together before marriage. One plausible explanation of this difference is that for noncohabiting couples the solidifying event in the relationship is the marriage itself. In contrast, for the cohabitors, marriage merely preserves the status quo and it is not until the event of a first birth that a significant change occurs. That is, for the noncohabiting couple, a first birth does not affect dissolution rates because the major structural change in the relationship occurs at the time of marriage when the couple begins to live together. However, for the cohabiting couple, the comparable change and cementing of the relationship does not take place until the first child is born. It is worthwhile to note that some cohabiting couples see having a child as an expression of their commitment and wish to marry before having that child.

Translation of the underlying hazard rates and proportionality factors into cumulative dissolu-

---

[9] We explored the possibility that the interpretations of results in Tables 4 and 6 are confounded by the fact that (a) long-term cohabitors are more likely to have premarital births; and (b) long-term cohabitors marry relatively late in life. We did so by estimating models that included relevant interaction terms. Inclusion of these terms did not contribute significantly to the explanatory power of the models.

Table 7. Monthly Hazard of Marital Dissolution and Covariate Effects—Only Women who Did Not Cohabit Premaritally

Covariate	Parameter Antilog (t-statistic)
Age at marriage <21 years	1.742 (3.44)
Premarital birth	1.578 (2.25)
First marital birth	.970* (0.12)
Occupation of main breadwinner during childhood = salaried employee	1.436 (2.08)

Duration (in years)	Monthly Hazard (λ) (t-statistic)
0–1	.000326 (24.3)
2–4	.000702 (26.0)
5–7	.000513 (23.2)
8–11	.000761 (23.5)
≥12	.000901 (23.4)

Number of observations = 969
log likelihood = − 1305.476

* Estimate not significantly different from zero at the .05 level.

tion probabilities yields statistics that allow one to see, in straightforward fashion, the vast differences in marital dissolution across various subgroups. We illustrate in Figure 2 the probability that a woman will have separated by selected durations of marriage. The range of results is startling. For example, a woman who is childless, has delayed marriage, and did not cohabit premaritally has a .08 probability of separating within 12 years of marriage. In contrast, her cohabiting counterpart is twice as likely to separate with a .16 probability. In addition, a woman who cohabited premaritally for more than three years, had a premarital birth, and married before she turned 21 years of age had a .54 probability of separating from her husband within her first 12 years of marriage.

SUMMARY AND CONCLUSIONS

The results of this paper indicate that Swedish women who cohabited with their first husband immediately prior to their marriage dissolve their first marriage at a significantly higher rate than married women who did not cohabit. Although the evidence is too indirect and fragmentary to pinpoint the precise cause of

higher marital dissolution rates among cohabitors, the findings are consistent with the hypothesis that those who cohabit may be a select group who tend to lack either the "interests" or "values" that are typically associated with marriage (Bernard 1982, p. 159).

The findings also indicate that the difference in dissolution rates between cohabitors and noncohabitors decreases as marital duration increases. We tested the hypothesis that diversity among cohabitors in the length of premarital cohabitation is partly responsible for the observed pattern of duration dependence. We found that among those who cohabited, women who lived with their future spouse for more than three years were significantly more likely to separate than those who cohabited for three years or less. This difference between long- and short-term cohabitors may reflect differences in the motivation to cohabit or in the extent to which patterns of individualistic behavior developed during the cohabitation period continue after marriage.
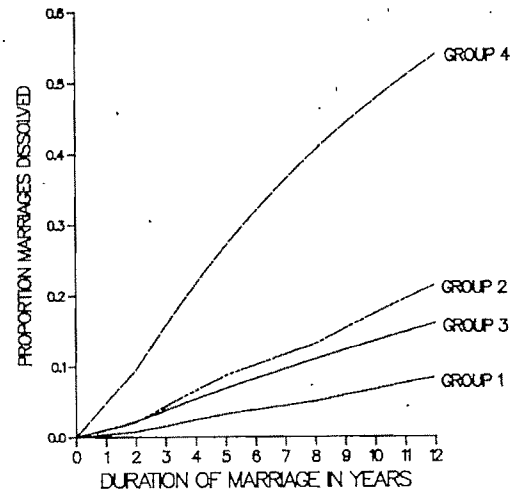
Fig. 2. Estimated Proportion of First Marriages Dissolved by Duration x for Selected Groups

Legend:
Group 1 = Did not cohabit, age at marriage 21 or greater, no premarital birth, no marital birth, main breadwinner's occupation during respondent's childhood not salaried employee.
Group 2 = Did not cohabit, age at marriage less than 21, premarital birth, no marital birth, main breadwinner's occupation during respondent's childhood not salaried employee.
Group 3 = Did cohabit for three years or less, age at marriage 21 or greater, no premarital birth, no marital birth, main breadwinner's occupation during respondent's childhood not salaried employee.
Group 4 = Did cohabit for greater than three years, age at marriage less than 21, premarital birth, no marital birth, main breadwinner's occupation during respondent's childhood not salaried employee.

Although strict inferences from the results reported here can only apply to the Swedish population, Balakrishnan's et al. (1987) analysis for Canada suggests that the nature, if not necessarily the magnitude, of the relationship between premarital cohabitation and marital stability is similar in other Western societies.

Blumstein and Schwartz's (1983) study of couples in the United States shows that cohabiting couples are more committed to personal independence than are married couples. This commitment is reflected in a lower likelihood of pooling income, owning joint property, and sharing leisure activities. Cohabitors do not expect the man to assume the role of provider; they expect partners to be responsible for their individual economic welfare. Furthermore, cohabiting couples are less likely than married couples to think monogamy is important and are more likely to approve of sex without love. Tanfer (1987), in the National Survey of Unmarried Women (a 1983 nationwide sample of never-married, 20- to 29-year-old women), found differences between those who had ever cohabited and those who had not. For example, those who had cohabited prior to the time of the survey attended church much less frequently and also were less likely at age 15 to be in households in which the parents were in intact marriages. These differences suggest that the direction of effects found in Sweden in all likelihood holds true in the United States as well.

In conclusion, simple descriptive statistics suggest no relationship between premarital cohabitation and subsequent marital stability. However, by applying a more refined model of marital dissolution, we have found the two events to be strongly negatively associated. This relationship is quite robust to varied model specifications. Due to limitations of the data, we cannot conclusively determine the mechanisms underlying this relationship. Nevertheless, the weight of the evidence does suggest that the higher marital-dissolution rates of cohabitors reflect their weaker commitment to the institution of marriage. Further insight into the nature and strength of the underlying structural relationships between premarital cohabitation and marital stability must await the development of richer data sets, especially those with more information on attitudes toward marriage.

REFERENCES

Audirac, Pierre-Alain. 1982. "Cohabitation et Mariage: Qui Vit avec Qui? (Cohabitation and Marriage: Who Lives with Whom?)" *Economie et Statistique* 145:41–59.

Balakrishnan, T.R., K.V. Rao, Evelyne Lapierre-Adamcyk, and Karol J. Krotki. 1987. "A Hazard Model Analysis of the Covariates of Marriage Dissolution in Canada." *Demography* 24:395–406.

Becker, Gary S., Elisabeth M. Landes, and Robert T. Michael. 1977. "An Economic Analysis of Marital Instability." *Journal of Political Economy* 85:1141–87.

Bernard, Jessie. 1982. *The Future of Marriage*. New Haven: Yale University Press.

Bernhardt, Eva and Britta Hoem. 1985. "Cohabitation and Social Background: Trends Observed for Swedish Women Born between 1936 and 1960." *European Journal of Population* 1:375–95.

Blanc, Ann Klimas. 1985. "The Effect of Nonmarital Cohabitation on Family Formation and Dissolution: A Comparative Analysis of Sweden and Norway." Ph.D. diss., Princeton University.

Blumstein, Phillip and Pepper Schwartz. 1983. *American Couples: Money, Work, Sex*. New York: William Morrow.

Brown, A. and K. Kiernan. 1981. "Cohabitation in Great Britain: Evidence from the General Household Survey." *Population Trends* 25:1–8.

Carlson, Elwood. 1986. "Couples without Children: Premarital Cohabitation in France." In *Contemporary Marriage: Comparative Perspectives of a Changing Institution*, edited by Kingsley Davis with A. Grossbard-Schechtman. New York: Russell Sage Foundation.

Cherlin, Andrew. 1977. "The Effect of Children on Marital Dissolution." *Demography* 14:265–72.

———. 1981. *Marriage, Divorce, Remarriage*. Cambridge: Harvard University Press.

Cox, D.R., and D. Oakes. 1984. *Analysis of Survival Data*. London: Chapman and Hall.

Davis, Kingsley. 1983. "The Future of Marriage." *Bulletin—The Academy of Arts and Sciences* 36:15–43.

DeMaris, Alfred and Gerald R. Leslie. 1984. "Cohabitation with the Future Spouse: Its Influence upon Marital Satisfaction and Communication." *Journal of Marriage and the Family* 46:77–84.

Festy, P. 1980. "Aspects Demographiques de la Formation de la Famille en Europe Occidentale (Demographic Aspects of Family Formation in Western Europe)." In *Marriage and Cohabitation in Contemporary Societies*, edited by J. Eekelaar and S. Katz. Toronto: Butterworth.

Glick, Paul C. 1984. "Marriage, Divorce, and Living Arrangements: Prospective Changes." *Journal of Family Issues* 5:7–26.

Hoem, Jan M. 1986. "The Impact of Education on Modern Family-Union Initiation." *European Journal of Population* 2:113–33.

Macklin, Eleanor D. 1978. "Nonmarital Heterosexual Cohabitation." *Marriage and Family Review* 1:1–12.

Mead, Margaret. 1966. "Marriage in Two Steps." *Redbook* (July):48–49, 84–85.

Menken, Jane, James Trussell, Debra Stempel, and Ozer Babakol. 1981. "Proportional Hazards Life Table Models: An Illustrative Analysis of Sociodemographic Influences on Marriage Dissolution in the United States." *Demography* 18:181–200.

Morgan, S. Philip and Ronald R. Rindfuss. 1985. "Marital Disruption: Structural and Temporal Dimensions." *American Journal of Sociology* 90:1055–77.

Murphy, M.J. 1985. "Demographic and Socio-economic Influences on Recent British Marital Breakdown Patterns." *Population Studies* 39:441–60.

Norton, Arthur J. 1983. "Keeping Up with Households." *American Demographics* (February):17–21.

Rosenblatt, Paul C. and Linda G. Budd. 1975.
"Territoriality and Privacy in Married and Unmarried
Cohabiting Couples." *Journal of Social Psychology*
97:67–75.
Spanier, Graham B. 1986. "Cohabitation in the 1980s:
Recent Changes in the United States." In *Contemporary Marriage: Comparative Perspectives of a Changing Institution*, edited by Kingsley Davis with A.
Grossbard-Schechtman. New York: Russell Sage
Foundation.
Statistics Sweden. 1981. "Women in Sweden." Stockholm. Inquiries about this data set should be addressed
to Statistics Sweden, S-115 81, Stockholm, Sweden.
Tanfer, Koray. 1987. "Patterns of Premarital Cohabitation among Never-Married Women in the United
States." *Journal of Marriage and the Family*
49:483–97.
Teachman, Jay D. 1982. "Methodological Issues in the

Analysis of Family Formation and Dissolution."
Journal of Marriage and the Family 44:1037–53.
Thornton, Arland. 1977. "Children and Marital Stability." *Journal of Marriage and the Family* 39:531–40.
Trost, J. 1980. "Cohabitation without Marriage in
Sweden." In *Marriage and Cohabitation in Contemporary Societies*, edited by J. Eekelaar and S. Katz.
Toronto: Butterworth.
Tuma, Nancy B. 1979. *Invoking RATE*, 2nd ed. Menlo
Park, CA: SRI International.
Waite, Linda J., Gus W. Haggstrom, and David E.
Kanouse. 1985. "The Consequences of Parenthood for
the Marital Stability of Young Adults." *American
Sociological Review* 50:850–57.
Westoff, Charles F. 1978. "Some Speculations on the
Future of Marriage and Fertility." *Family Planning
Perspectives* 10:79–83.

SCHOLARLY CONSENSUS AND JOURNAL REJECTION RATES*

LOWELL L. HARGENS

University of Illinois at Urbana-Champaign

Rejection rates for scholarly journals show substantial variation between disciplines. Explanations of this variation have focused on two possible sources: variation in consensus and in space shortages. Longitudinal data on journal rejection rates show that they have been very stable over time and are largely unaffected by changes in submissions, impugning the argument that space shortages explain disciplinary variation in rejection rates. In contrast, a model of the manuscript-evaluation process can account for the observed variation in rejection rates and also casts light on additional characteristics of manuscript evaluation processes in different disciplines as well. Possible links between consensus and each of the elements of the model are discussed.

Scientific journals play a dual role in scientific communities. They are both a means by which a community certifies additions to its body of accepted knowledge and a means through which individual scientists compete for priority and recognition (Merton 1957; Hagstrom 1965; Ziman 1968). Although several studies have examined the second role (Crane 1967; Zuckerman and Merton 1971; Pfeffer et al. 1977), only a few researchers have studied the processes by which scientific journals select manuscripts for publication. The most important of these is Zuckerman and Merton's (1971) study, which, among other topics, examined disciplinary variation in rejection rates. They found substantial variation, with rejection rates of 20 to 40 percent in the physical sciences, and 70 to 90 percent in the social sciences and humanities.

Zuckerman and Merton suggested two possible sources for the wide variation in rejection rates: variation in consensus across fields of scholarship and differences in space shortages. As evidence for the former, they showed that substantial differences in rejection rates also exist *within* hybrid fields. For example, cultural anthropology journals have higher rejection rates than physical anthropology journals, and rates for journals in social, abnormal, and educational psychology exceed those in experimental, comparative, and physiological psychology. Zuckerman and Merton also reported

impressionistic evidence that other features of the review process, such as decision rules that minimize ill-advised acceptances rather than ill-advised rejections, vary systematically across fields. As evidence for the claim that space shortages may cause the interdisciplinary variation in rejection rates, they noted that, as of the late 1960s, space shortages in the social sciences may have exceeded those in the physical sciences because the number of social scientists was growing faster than the number of physical scientists.

Few researchers have pursued Zuckerman and Merton's provocative study, in part because in recent years they have tended to emphasize similarities across scholarly fields rather than differences between them (Whitley 1984, pp. 3–7). For example, Cole (1983) argues that disciplinary elites acting as gatekeepers and evaluators produce and maintain consensus on appropriate research questions and techniques. Cole claims that all scholarly fields face similar functional problems and, therefore, have similar elite structures, and that this, in turn, produces similar levels of consensus at the "research frontiers" of different disciplines. Knorr-Cetina (1981) reaches a similar conclusion from different premises. She argues that processes of knowledge production in the natural and social sciences are similar because they consist of research practices that are contextually contingent and depend largely on social negotiation. The substantial variation in journal rejection rates Zuckerman and Merton found seems inconsistent with both of these portrayals, and arguments for the similarity of research work across fields usually either ignore the rejection rate data or dismiss them as resulting from differential space shortages (Cole 1978; 1983).

In this paper, I examine the causes of disciplinary variation in rejection rates more closely. The discussion has three parts. First, I present data that cast doubt on the space-

shortage explanation of differences in rejection rates. Second, I develop and evaluate a formal model of the manuscript-refereeing process. The model shows that three characteristics of that process can account for the observed variation in rejection rates, and that they also elucidate other aspects of the procedures by which journals evaluate manuscripts. Third, I attempt to specify more precisely the mechanisms through which variation in consensus may influence the manuscript-refereeing process. The results indicate that rejection rates may be a relatively good measure of levels of consensus, and, therefore, that levels of consensus vary substantially across fields.

REJECTION RATES AND SPACE SHORTAGES

Testing the space-shortage argument with cross-sectional data is difficult because direct measures of journal space shortages are not available, and most indirect measures are probably contaminated by other variables, such as levels of competition for priority (Hargens 1975, p. 21). In addition, it is impossible to determine if the fact that social scientists were proliferating faster than physical scientists in the late 1960s led to greater space shortages in the former fields because we have no comparable data on respective increases in publication outlets at that time.[1] Thus, to assess the space-shortage argument, I collected longitudinal data on the number of papers submitted to and published by 30 U.S. scholarly journals from 1967 to 1983. The journals included are published by the major scholarly associations in physics, chemistry, biology, psychology, anthropology, sociology, economics, and political science.[2] I began by estimating each journal's

rejection rate in the late 1960s by dividing the number of papers it published during 1968–70 by the number of submissions received during 1967–69 and subtracting this ratio from unity.[3] I constructed similar estimates for the period 1980–83. To measure changes in submissions, I divided a journal's 1981–82 submissions by its 1967–69 submissions. On average, submissions to these journals increased by 30 percent, although six journals had increases of over 50 percent and six experienced decreases. One can determine the extent to which changes in submissions explain changes in rejection rates by regressing the early 1980s rejection rates on the late 1960s rejection rates and the measure of the change in submissions. Table 1 shows the results of this analysis.

Although changes in journals' submissions varied considerably, their independent effect on the 1980s rejection rates is very small. In contrast, the independent effect of the earlier rejection rates is very strong, indicating considerable stability in the journals' rejection rates.[4] This result, in turn, impugns the claim that space shortages resulting from increased submissions can account for the substantial disciplinary differences observed in journal rejection rates.[5]

---

[1] Collecting adequate data on this question would be extremely difficult because one would have to take into account the expansion of existing journals, as well as the establishment of new ones, and also confront the thorny problem of what journals are included in what fields.

[2] The names of the journals and the data summarized below are presented in the Appendix. These journals tend to be among the most prestigious in their fields, and as a group are similar to those Zuckerman and Merton studied (1971, p. 75). The extent to which the results of the analyses reported below also hold for less prestigious journals and those published overseas is unknown. For example, because British journals are somewhat less likely to use external referees than U.S. journals, and British editors have more autonomy than their U.S. counterparts (Whitley 1970; Gordon 1978), my analyses of elements of the refereeing process may not describe British journals as well as they do U.S. journals.

There is little information on whether manuscript evaluation processes vary for journals of varying prestige, but data on journal prestige and rejection rates

for 42 of the 63 sociological journals Glenn (1971) ranked indicate relatively little variation in the rejection rates, which are, in turn, very weakly correlated to journal prestige. Specifically, data from Mullins (1977) and Huber (1982) for the journals show a mean rejection rate of 83 percent with a standard deviation of 7 percent and a correlation between journal prestige and rejection rates of only .09. Similarly, data on 31 economics journals whose rejection rates are reported in Adams (1977) or Oster (1980) and whose prestige ratings are reported in Hawkins et al. (1973) show a mean rejection rate of 84 percent, a standard deviation of 6 percent, and a correlation between prestige ratings and rejection rates of .04. Thus, as far as sociology and economics are concerned, rejection rates observed for high-prestige journals resemble those for the whole field.

[3] Three-year periods were used in computing these ratios to minimize unreliability due to annual variation in numbers of submissions and articles published. Such variation is more marked among social science journals because they are relatively small. I used a one-year lag in constructing the rejection rate estimates because only annual data are available and both no lag and a two-year lag seem unrealistic (King et al. 1981, p. 201). In fact, however, annual variation in numbers of submissions and publications is so small that the results of analyses using different lags are nearly identical to those reported below.

[4] Data for a few journals for a portion of the entire period also show comparatively little variation in yearly rejection rates. For example, PMLA shows a rejection rate of 94 percent during 1973–75 and 95 percent during 1980–82 (Showalter 1984).

[5] The impressive stability in the rejection rates also casts doubt on another factor often advanced to explain physical science journals' relatively low rejection rates: page charges. For example, the Physical Review has

Table 1. Regression of 1981–83 Journal Rejection Rates on 1967–69 Rejection Rates and Growth in Submissions, for 30 U.S. Scholarly Journals

Variable	Metric Coefficient	Standardized Coefficient	t
Growth in submissions	.056	.08	1.29
1967–69 rejection rate	1.115	.95	15.26
Constant	−.094	—	−1.27
$R^2 = \quad .90$			
$\bar{R}^2 = \quad .89$			
$F_{2,27} = 116.72$			
Correlation			
	2	3	
1. 81–83 rejection rate	.943	−.047	
2. 67–69 rejection rate	—	−.134	
3. Growth in submissions		—	

A credible explanation for these results can be found in editors' reports that sometimes accompany the publication of information on submissions and acceptances. First, it is noteworthy that editors of journals with high rejection rates view those rates as "normal." Indeed, they often complain of a dearth of publishable papers among submissions. Editors' reports usually focus on backlogs of accepted papers. They suggest that referees' judgments produce a crop of publishable papers each year, and, when the crop is too large to be published during a year, a backlog accumulates. If this occurs for several years the backlog becomes unacceptably large, and ameliorative action may be taken. Not surprisingly, editors overwhelmingly favor expanding their journals rather than raising their rejection rates. In some instances, however, this last-ditch measure must be taken. For example, during the early and mid-1970s, the *American Political Science Review* built up a backlog of almost two years of accepted papers. In addition to expanding the journal, the editor reduced the acceptance rate for one year to six percent, nearly half its normal rate (Jones 1977; 1978). But, in general, editors rarely report that they have been forced to reject more papers than the reviewing process dictates. When they must do so, it is for a fairly short period.

Thus, the editors' reports suggest that space shortages affect journals' backlogs rather than their rejection rates, and that, even when the latter are affected, they tend to return to normal relatively quickly. This explanation accounts for the long-term stability in journal rejection rates, even in the face of substantial changes in submissions. Of course, editors may wish to convey the impression that they play a largely

unobtrusive role in the peer-review process, and their annual reports might contain disingenuous elements. However, given their consistency with the above results, it seems reasonable to conclude that space shortages cannot explain disciplinary variation in journal rejection rates, and, instead, to look to the peer-review process itself as a possible source of such variation.

A MODEL OF THE PEER REVIEW PROCESS

In order to analyze the operation of journal peer-review processes, I have constructed a formal model that identifies three basic variables and shows how they jointly produce various outcomes. This model is an analytic device; it does not attempt to provide a comprehensive view of all the variables operating in the review process. Instead, it is a "thought experiment" that employs ideal types and holds constant complicating variables whose effects often obscure the relations under study in this paper (for a good exposition of this analytic strategy, see Keyfitz 1975). My model includes two variables noted in past work: a journal's "decision structure," and the correlation between referees' evaluations of manuscripts. Zuckerman and Merton (1971) note that journals with high acceptance rates tend to use procedures that presuppose that submitted papers should be published and to minimize the chance that worthy papers will be rejected, while the opposite holds for journals with low acceptance rates. These different presuppositions give rise to different reviewing structures that are distinguished by the number of referees to whom a manuscript is sent initially. In the first, "decision structure 1," a manuscript is initially sent to only one referee.[6] If the referee

---

substantially reduced its page charges since the late 1970s, but rejection rates and submissions have remained fairly stable. Editors of such journals have told me that they view page charges as an indirect subsidy to subscribers rather than as a source of revenue to increase the number of papers they publish.

[6] My depiction of this decision structure draws upon informal discussions with editors of physical science journals, and descriptions in Goudsmit (1969) and Zuckerman and Merton (1971).

recommends acceptance, the editor accepts it. If the referee recommends rejection, the editor usually seeks the opinion of a second referee to guard against the possibility that the initial referee's opinion is idiosyncratic and unrelated to the paper's merit. A negative evaluation from the second referee is usually sufficient to prompt the editor to reject it. On the other hand, a favorable recommendation from the second referee suggests that the paper is "problematic"—neither clearly acceptable or unacceptable. A problematic paper is likely to be sent to a third referee, whose opinion will decide its fate for that journal.

In contrast, "decision structure 2" employs two initial referees for each manuscript, and requires that both recommend acceptance for immediate acceptance by the editor. If both referees recommend rejection, the editor rejects the paper. A paper that receives a split decision from the initial referees is sent to a third, who determines its fate. Panel A of Table 2 represents these two decision structures schematically.

As models, either perfectly reflects the actual operation of a given journal. Neither model assigns the journal editor a role beyond selecting referees and compiling their "votes," while, in reality, editors sometimes differentially weight the recommendations of individual reviewers. Also, if one referee spots a fatal flaw in a manuscript, the editor may reject it without a third review even though a second referee returns a favorable recommendation. Finally, when referees' reviews are apparently neither positive or negative, an editor may seek fourth and even fifth opinions on a manuscript. Thus, neither of the two ideal-typical decision structures in Table 2 is a perfect description of a journal's manuscript review process. Nevertheless, it is instructive to compare how each affects various features of that process. Doing so allows us to isolate these effects from the influence of factors, such as those enumerated above, that result from actions of referees and editors not encompassed by the two structures.

A second variable in the peer-review process is the level of agreement among reviewers. Most studies of this topic have focused on whether referees' assessments of a manuscript's merit are reliable and have examined the correlation between referees' recommendations to measure inter-rater reliability. Most of these studies have examined behavioral science journals, and have

Table 2. Two Models of Journal Peer-Review Procedures

A. The Models and Their Constituent Outcomes

	Decision Structure 1					Decision Structure 2			
Outcome Category	Referee's Recommendation			Editor's Decision	Outcome Category	Referee's Recommendation			Editor's Decision
	1st	2nd	3rd			1st	2nd	3rd	
A	a	—	—	a	A	a	a	—	a
B	r	a	a	a	B	a	r	a	a
C	r	a	r	r	C	r	a	a	a
D	r	r	—	r	D	a	r	r	r
					E	r	a	r	r
					F	r	r	—	r

B. Proportion of Manuscripts Falling into Each Outcome Category in Panel A

	Referee's Recommendations	
Decision Structure 1 Outcome	Statistically Independent	Perfectly Correlated
A	P	P
B	$P^2 Q$	Null
C	$P Q^2$	Null
D	Q^2	Q
	1.0	1.0
Decision Structure 2 Outcome Category		
A	P^2	P
B	$P^2 Q$	Null
C	$P^2 Q$	Null
D	$P Q^2$	Null
E	$P Q^2$	Null
F	Q^2	Q
	1.0	1.0

Note: In Panel A, a = accept and r = reject. In Panel B, P = proportion of referees' reports that recommend acceptance and $Q = 1 - P$.

found levels of referee agreement that are uncomfortably close to statistical independence (for a discussion of these results, see Peters and Ceci 1982, and the commentary on Peters and Ceci in the same issue). One of the few such studies for physical science journals found high referee agreement (Zuckerman and Merton 1971).[7] Thus, in the models below, I specify statistical independence and perfect agreement as two extremes to determine the effects of variation in referee agreement on outcomes of the peer-review process.

In addition to decision structure and referee agreement, models of variation in the review process must incorporate a third variable, the proportion of referees' recommendations that favor publication. This proportion is obviously higher for journals with low rejection rates than for those with high rejection rates, but there has been no assessment of its range of values across journals. From data reported in the few studies that contain sufficient information, I estimate that the proportion varies between .2 and .7 across journals,[8] and use these extremes in the analyses below. Given information on each of the three variables discussed above, one can derive various outcomes of the manuscript evaluation process, such as rejection rates, initial manuscript dispositions, lags between submission and publication, etc. I turn first to journal rejection rates.

Journal Rejection Rates

Panel B of Table 2 presents expressions for the proportion of manuscripts that fall into each of the possible outcome categories for the two decision structures in Panel A. For example, letting P represent the proportion of reviews that are favorable to publication, the proportion of manuscripts that receive one positive and two negative recommendations under decision structure 1 (Outcome C) will equal PQ^2 when referees' recommendations are statistically inde-

Table 3. Rejection Rates Generated by the Two Decision Structures, with Varying Values of P and Varying Correlation Between Referees' Recommendations

Referees' Recommendations	Decision Structure 1		Decision Structure 2	
	$P = .7$	$P = .2$	$P = .7$	$P = .2$
Statistically independent	15%	77%	22%	90%
Perfectly correlated	30%	80%	30%	80%

pendent ($Q = 1 - P$). Since Outcomes C and D of decision structure 1 and Outcomes D, E, and F of decision structure 2 result in rejection of manuscripts submitted for publication, it is possible to determine the journal rejection rates that would be generated by each combination of values for the three variables discussed above. These rejection rates are presented in Table 3.

The most important feature of these results is that variation in the three variables of the peer-review process produces a wide range of rejection rates—from 15 percent to 90 percent. Since these figures encompass the range of rejection rates that have been observed in natural and social science journals, one may conclude that variation in these three variables alone can explain the observed variation in rejection rates.

The results in Table 3 also show that the proportion of reviews that recommend publication is the most influential of the three variables, but that the three variables have an interaction effect on rejection rates. Let us examine the effects of journal decision structure and referee agreement in more detail, beginning with the influence of the former. The difference in rejection rates produced by the two structures equals the proportion of manuscripts that receive an acceptance from the first referee and rejections from referees 2 and 3 (denoted as Outcome D in decision structure 2). Manuscripts that fall into this category are rejected under the second decision structure but would have been accepted under the first. When referees' recommendations are perfectly correlated, no manuscripts fall into this category: the proportion increases to PQ^2 when recommendations are statistically independent. In the latter case, the proportion reaches a maximum value of .148 when $P = 1/3$.[9] Thus, the maximum effect of variation in decision structure under the range of conditions considered here is substantially smaller than the effect of variation in P.

Table 2 also shows that variation in the correla-

---

[7] One possible reason for the smaller number of studies of referee agreement in physical science journals is the fact that such journals are more likely to use the first decision structure. As a result, only a nonrepresentative subset of all manuscripts receive two reviews. Thus, the findings of such studies may not reflect the level of agreement that would occur if all papers received two reviews.

[8] Between 1958 and 1961, 36 percent of referees' reports to *Social Problems* recommended publication (Smigel and Ross (1970). Unpublished work by Bakanic (1985) indicates that this proportion was 24 percent for the *American Sociological Review* during 1978–81. In contrast, the data reported by Zuckerman and Merton (1971, p. 67) imply that 67 percent of referees' recommendations to the *Physical Review* favored publication.

[9] This value of P is determined by setting the first derivative of the function $f(P) = P(1 - P)^2$ to zero and solving for P. The same procedure is used for determining values of P in other cases reported below.

tion between referees' recommendations affects rejection rates by determining the proportion of manuscripts that receive mixed recommendations (Outcomes B and C under the first decision structure and B through E under the second). When referees' recommendations are perfectly correlated, no manuscripts receive mixed recommendations; as referees' recommendations approach statistical independence, an increasing proportion receive mixed recommendations. The two decision structures have different effects on rejection rates when a large proportion of manuscripts received mixed recommendations (see Figure 1).

Under the first decision structure, lower consensus among referees always *decreases* a journal's rejection rate. This can be seen by noting first that the proportion of manuscripts accepted under this decision structure is the sum of the proportions that fall into outcome categories A and B, and that the proportion that fall into category A equals P regardless of the overall correlation between referees' recommendations. Thus, variation in referee consensus affects the rejection rate of a journal using decision structure 1 only through its effect on the proportion of manuscripts that fall into outcome category B. This proportion increases from zero, when referees' recommendations are perfectly correlated, to P^2Q when they are statistically independent (and P does not equal either zero or unity). In the latter case, the proportion falling into category B can be as large as .148, which occurs when $P = 2/3$. Thus, other things equal, lower referee consensus decreases the overall rejection rate of journals using the first decision structure—this is one way in which that structure gives submitted papers the benefit of a doubt.

The second decision structure reacts to independent referee recommendations by increasing the rejection rate when the majority of recommendations are unfavorable and by reducing it when the majority are favorable. This is easiest to see by comparing the acceptance rate when recommendations are perfectly correlated (P) with the rate when they are statistically independent ($P^2 + 2P^2Q$) (see Figure 1). When P equals 0, .5 or 1.0, P equals $P^2 + 2P^2Q$ and it makes no difference whether recommendations are perfectly correlated or independent. When $0 < P < .5$, however, P is greater than ($P^2 + P^2Q$) and statistically independent recommendations produce a lower acceptance rate than perfectly correlated recommendations. When $.5 < P < 1.0$, independent recommendations produce higher acceptance rates than perfectly correlated recommendations.[10]

In sum, variation in the three variables analyzed above can produce the observed variation in journal rejection rates. The three variables have a complicated interaction effect on rejection rates in which the highest rejection rates are produced by the combination of low values of P, decision structure 2, and statistically independent referee recommendations, while the lowest rejection rates are produced by the combination of high values of P, decision structure 1, and statistically independent recommendations. Of the three variables in this interaction, however, variation in P is the most influential.

Initial Manuscript Dispositions

The fact that the three variables analyzed above can account for observed variation in journal rejection rates does not necessarily imply that they do, in fact, produce that variation. As noted above, in these models editors play only a "recordkeeping" role, while, in reality, they are free to discount or ignore referees' recommendations. To the extent that they do, variation in rejection rates will be due to editors' behaviors omitted from the above models.

Assessing this possibility requires data on the relation between referees' recommendations and editors' decisions. Fragmentary data for two journals permit a preliminary test of the adequacy of the above analyses. The data consist of editors' initial dispositions of cohorts of manuscripts submitted to the *Physical Review* (PR) and the *American Sociological Review* (ASR), and, for the latter, the initial referees' recommendations. By "editors' initial dispositions," I mean whether editors accept, reject, or seek revisions and/or further opinions about a manuscript. "Initial referees' recommendations" refers to whether the initial referees of a manuscript recommend that it be accepted,
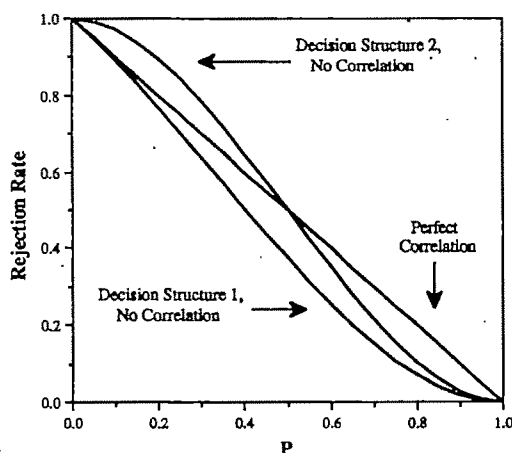
Fig. 1. Rejection Rates Generated by Different Decision Structures, Varying Correlations between Referee Recommendations, and Varying Values of P

Table 4. Dispositions of Manuscripts Generated by Varying Decision Structures, Values of P and Correlations between Referees' Recommendations, and Observed Dispositions for Two Journals

A. Dispositions Predicted by Models

Disposition	Statistically Independent Referee Recommendations				Perfectly Correlated Referee Recommendations	
	Decision Structure 1		Decision Structure 2			
	$P = .67$	$P = .24$	$P = .67$	$P = .24$	$P = .67$	$P = .24$
Immediately accepted	67%	24%	45%	6%	67%	24%
Problematic	22	18	44	36	0	0
Immediately rejected	11	58	11	58	33	76
	100%	100%	100%	100%	100%	100%

B. Observed Dispositions for ASR and PR

	ASR^a Referees		ASR^a Editor		PR^b Editor
Both recommend acceptance	10%	Immediate acceptance	11%	Immediate acceptance	65%
Split	33	Revise and resubmit	24	"Problematic"	23
Both recommend rejection	57	Immediate rejection	65	Immediate rejection	12
	100%		100%		100%
N	(342)		(342)		(approx. 7000)

a Source: Bakanic 1986.
b Source: Zuckerman and Merton 1971, p. 91.

rejected, or hand down a split decision.[11] Since the models above offer specific predictions about the proportions of manuscripts that fall into each of the initial referee recommendation categories, it is possible to assess the adequacy of emphasizing referee recommendations by comparing those predictions with observed distributions of editors' decisions.

Panel A of Table 4 presents the predicted proportions of manuscripts that would be immediately accepted, immediately rejected, or for which revisions and/or further opinions would be sought, under the range of conditions studied in the above analyses. Specifically, the manuscripts that fall into Outcome A for each decision rule would be immediately accepted, those that fall into Outcome D for decision structure 1 and Outcome F for decision structure 2 would be immediately rejected, and those that fall into the intermediate cases for each decision structure would elicit revisions and/or further opinions. The first four columns of Panel A present results for combinations of values of P and type of decision structure for the case where

referees' recommendations are statistically independent. Since type of decision structure has no effect when referees' recommendations are perfectly correlated, results for this case are presented in the last two columns of Panel A. The values of P that generate the distributions of manuscript dispositions in Panel A are those reported above for ASR and PR: .24 and .76 respectively. Panel B of Table 4 presents data on the observed dispositions of manuscripts for ASR and PR. The first column shows the recommendations of the initial referees of ASR submissions, and the second and third columns show the editors' initial dispositions of papers submitted to ASR and PR respectively.

Let us first examine the correspondence between referees' recommendations and the predictions of the models presented above. The distribution of ASR referees' recommendations strongly resembles what decision structure 2 with $P = .24$ and statistically independent referees' recommendations would yield. The close correspondence of these predictions with the observed percentages in the first column of Panel B indicates that ASR referees' recommendations, like those for other behavioral science journals, are nearly statistically independent.[12]

[11] "Initial referees" consist of the first two referees in all cases except Case A, decision structure 1, which has only one initial referee. In the analysis below, I categorized recommendations that a manuscript be accepted or accepted conditional on minor revisions as recommendations that a paper be accepted, and all others—revise and resubmit, submit to another journal, and reject outright—as recommendations that it be rejected.

[12] Cohen's kappa, a PRE measure of rater agreement (Cohen 1960), is a modest .15 for these data. However, the null hypothesis that the referees' recommendations are statistically independent can be rejected with some confidence (chi-squared equals 7.34 with 1 d.f.). It is worth noting that levels of referee agreement for the ASR

Comparing the first and second columns of Panel B yields information on whether initial editorial dispositions for *ASR* correspond to initial referees' recommendations. The strong similarity of these two distributions suggests that referees' recommendations are the major factor determining the editors' dispositions.[13] Thus, for *ASR* at least, it appears that the three variables in the manuscript-reviewing process discussed above largely account for the observed outcomes of that process.

Data for the initial editorial dispositions of manuscripts submitted to *PR* in Panel B bear a close resemblance to what would occur for a journal using the first decision structure with $P = .67$ and statistically independent referee recommendations. Since data on initial referees' recommendations for these manuscripts are not available, I cannot verify the third condition (statistical independence), but the results for *PR* are clearly consistent with the model developed above.

Two additional features of the results in Table 4 deserve mention. First, the values shown in Panel A for the cases where referees' recommendations are perfectly correlated seem far off the mark. This is because all manuscripts would be immediately accepted or rejected if referees' recommendations were perfectly correlated and editors simply took their referees' advice. In fact, a substantial proportion of the manuscripts submitted to each journal receive neither of these two dispositions. Second, the choice of a decision structure strongly affects the distribution of initial referees' recommendations by affecting the number of referee disagreements, and, presumably, this also affects initial editorial dispositions. Although use of the second decision structure has only a small effect on the overall rejection rate (see Table 3), it produces substantially more split referee recommendations and probably leads editors to seek revisions and/or further opinions more often than the first. For example, Table 3 shows that, when $P = .2$ and referees' recommendations are independent, 23 percent of manuscripts are eventually accepted under the first decision structure, and since the proportion of manuscripts that are immediately accepted equals $P = .2$, 87 percent ($20\%/23\% = .87$) of all acceptances are immediate acceptances. In

contrast, under the same conditions, the second decision structure produces an acceptance rate of 10 percent, but only 40 percent ($4\%/10\%$) of these acceptances are immediate. Thus, the second decision structure requires more referee reports and revisions from authors, yielding a more costly review process.

Manuscript Evaluation Time Lags

The close fit between the observed distributions of manuscript dispositions and those predicted by the models presented above suggests that structural characteristics of the review process may explain other kinds of variation among scholarly journals. For example, it seems likely that the factors discussed above contribute to the considerably longer time lags between submission and publication in social science than in physical science journals found by Garvey, Lin, and Nelson (1970). Insofar as physical science journals are more likely to use decision structure 1 and have high values of P, large proportions of the papers they publish will have had a single referee. In contrast, insofar as social science journals use decision structure 2, all their published papers will have been reviewed at least twice. Clearly, the use of a second referee slows down the manuscript evaluation process simply because an editor must wait for two reviews. In addition, I noted above that use of decision structure 2 causes a substantial proportion of eventually published papers to be reviewed by more than two referees. Editors are probably likely to require that such papers be substantially revised before they are accepted for publication. Thus, the features of the manuscript review process examined above can account for differences in the average time lags between submission and editorial acceptance of manuscripts.

Zuckerman and Merton (1971) conjectured that journals with high rejection rates will dispose of eventually rejected manuscripts more rapidly than journals with low rates. It is clear that journals with low values of P will reject larger proportions of eventually rejected manuscripts after only two referees have returned their recommendations than will journals with high values of P. In addition, use of decision structure 1 means that the first two reports will be obtained seriatim rather than contemporaneously. As a result, journals that use decision structure 1 and have high P values will take longer, on the average, to reject papers than those that use decision structure 2 and have low P values.[14] These considerations imply an

---

and other social science journals are well below the reliability coefficient of .5 assumed by Stinchcombe and Ofshe (1969) in their model of the journal decision process. Thus, their analysis overstates the correlation between papers' probabilities of acceptance and "quality," however the latter may be socially defined.

[13] In fact, Bakanic's data show a gamma coefficient of .89 for the relation between referees' recommendations and editors' initial dispositions.

---

[14] In part, this pattern will result if editors of high rejection-rate journals are more likely to reject papers without outside review. Zuckerman and Merton cite

affinity between high values of P and the use of decision structure 1 on the one hand, and low values of P and the use of decision structure 2 on the other. Each combination tends to minimize the time required to arrive at its particular modal outcome.

CONSENSUS AND THE REVIEW PROCESS

The above analyses suggest that the three variables examined in this paper provide a useful conceptual model for understanding how scholarly journals evaluate manuscripts. This model can account for variation in rejection rates, the proportion of papers that must be revised and resubmitted for publication, and the lengths of various time lags. The three variables are only proximate determinants of such outcomes, however, and analyses of their own determinants are necessary. Let us follow Zuckerman and Merton's (1971) suggestion that disciplinary differences in rejection rates reflect differences in consensus, and inquire about the links that may exist between consensus and each of the three variables discussed above.[15]

P, the proportion of referees' reports that recommend a manuscript be published, reflects the overall level of agreement between referees and authors on whether manuscripts merit publication in a given journal.[16] Initially, the connection between such agreement and disciplinary consensus seems fairly straightforward and strong. When scholars do not share conceptions of appropriate research problems, theoretical approaches, or research techniques, they tend to view each other's work as deficient and unworthy of publication. Lack of consensus should, therefore, produce low values of P. It may be, however, that this link between

---

examples of editors rejecting 40 percent of submitted manuscripts without outside review because the manuscripts fall far short of minimum standards of scholarship.

[15] Whether there is significant variation in consensus among disciplines is a point of contention among sociologists (cf. Cole 1983; Hargens and Hagstrom 1982). Insofar as there are probable links between consensus and the three variables, the case for interdisciplinary variation in consensus is strengthened by the rejection rate data reported by Zuckerman and Merton (1971).

[16] One might object that this interpretation underestimates the role of journal editors in determining the eventual dispositions of manuscripts. Specifically, if editors were the real arbiters of whether papers are accepted and exert their influence by choosing referees who they know will agree with their opinions, then P would measure the extent of disagreement between authors and editors. However, the typically low correlations between referee recommendations seem to contradict this argument.

consensus and P is not their only causal connection. Figure 2 shows possible indirect causal links involving (1) the proportion of manuscripts that editors summarily reject; (2) the diffuseness of a field's journal system; and (3) editors' use of referees who are likely to disagree about the merit of a manuscript.

As indicated in Figure 2, the level of consensus in a field is probably negatively related to the proportion of submissions that editors reject without outside review (Zuckerman and Merton 1971). If editors can correctly anticipate referees' recommendations, the proportion of summary rejections should be positively related to P, since editors are weeding out manuscripts that referees would recommend rejecting. In contrast, if low levels of consensus reduce correlations between editors' and referees' assessments, the summary rejection of large proportions of submissions may have only a small effect on P. As long as editors' and referees' assessments are positively related, however, the link between consensus and P through the proportion of summary rejections will reduce the overall correlation between consensus and P below what would obtain if only consensus affected P. If editors summarily rejected no papers, differences in typical values of P between the social and physical sciences might be even larger than those currently observed.

Consensus may indirectly affect P in a second way: through the "diffuseness" of fields' systems of journals. In the physical sciences, relatively few journals publish a majority of the literature (a "concentrated" journal system, according to Garvey et al. 1970). In the social sciences, the largest journals publish small proportions of the literature (a "diffuse" system). It is reasonable to argue that the level of consensus in a field is negatively associated with the diffuseness of its journal structure: low consensus means more disagreement on the merit of submitted work, and such disagreement leads editors to complain that researchers are not producing enough publishable papers. In such fields, the few dominant journals, typically

Fig. 2. A Model of the Relations between Consensus and P

those published by the field's professional association, have little motivation to expand to the point where they publish most of the field's literature. In contrast, in fields where referees typically recommend publication, core journals may experience pressure to expand to accommodate the volume of publishable papers.

Conceivably, the diffuseness of a field's journal system may independently reduce the field's average value of P. An author may find it harder to discover the "right" journal for a manuscript in a field with a large number of small journals of varying substantive specialization and prestige. Thus, low consensus may lower the average level of P through its tendency to produce a diffuse journal system.

However, little evidence points to a negative relationship between diffuseness and P. If a diffuse journal system leads authors to submit their work to inappropriate journals, this should be a common reason for rejecting papers. Garvey et al. (1970) studied reasons given for rejecting papers in a number of fields and found that rejected manuscripts in the physical sciences, where journal structures are more concentrated, are *more* likely than those in the social sciences to be rejected because their subject matter is judged inappropriate. Moreover, social science manuscripts that receive this criticism usually receive other criticisms as well, suggesting that social science editors typically use this objection euphemistically (Garvey et al. 1970). In contrast, social science journals are much more likely than physical science journals to reject papers on methodological and theoretical grounds. These results not only fail to support the hypothesized negative relationship between diffuseness and P, they also strongly support the direct link between consensus and P shown in Figure 1.

The third indirect link between consensus and P shown in Figure 2 results from editors' strategies in selecting referees when they consult at least two initial referees. In low-consensus fields, editors often know that researchers in some areas disagree on theory and method, and may therefore solicit the opinions of referees on both sides of a disagreement. In high-consensus fields, scholars are likely to subscribe to a "positivist ideal" (Hagstrom 1965, pp. 255–91); editors in such fields are likely to seek "neutral" referees or even choose at random from among relevant specialists.[17]

Choosing referees with opposing perspectives will either increase or decrease the value of P, depending on whether P is otherwise small (less than .5) or large (greater than .5). When P is otherwise small, selecting opposing referees will tend to increase its value because one referee can be expected to be positively predisposed toward a manuscript. When P is otherwise large, this strategy will tend to decrease it because one referee will probably be negatively predisposed. Thus, using this strategy will reduce differences between high- and low-consensus fields in average values of P.

In sum, the indirect links shown in Figure 1 between consensus and P are probably weaker than the direct link. In addition, because the two negative indirect links that tend to reduce the relationship between consensus and P probably outweigh the positive indirect link through the diffuseness of the journal system, observed variation in P across disciplines may actually understate the variation that would exist if only the direct link operated. Thus, it is likely that interdisciplinary variation in P ranks fields in disciplinary consensus relatively reliably because of its close relation to the evaluation of recently completed research.

Whether a journal uses decision structure 1 or 2 appears to be indirectly linked to consensus through its correlation with P. The association between high values of P and decision structure 1 has at least two possible sources. The first, already noted, is that each decision structure tends to minimize the time required to reach the modal decisions that either high or low values of P produce. The second consists of a basic social-psychological principle: people tend to attribute greater validity to expected observations than to unexpected ones (Tversky and Kahneman 1974). Thus, editors of journals with high values of P probably see positive referee recommendations as more valid than negative ones. This may account for their willingness to accept a paper on the strength of a single positive recommendation and their unwillingness to reject a paper on the strength of a single negative one. In contrast, editors of journals with low values of P probably attribute greater validity to negative recommendations than to positive ones, and, therefore, require additional opinions to confirm unexpected positive recommendations. Note that both of these explanations imply that a decision structure in which a paper is rejected on the basis of a negative recommendation from a single initial referee would be "optimal" for the operation of journals with low values of P. However, authors'

---

[17] I thank an anonymous referee for bringing this indirect link to my attention. The referee also notes that choosing referees with opposing points of view tends to increase an editor's power in making decisions about manuscripts by providing referee recommendations to which an editor can appeal regardless of the decision. Use of this strategy also reduces the correlation between

referee recommendations. Finally, use of a single initial referee can be viewed as one form of commitment to the positivist ideal.

probable opposition to rejections based on a single review precludes its use.

The magnitude of P is certainly not the only determinant of journals' choices of decision structures. Although these two variables are moderately correlated for journals published under the auspices of major U.S. scientific associations, Gordon's (1978) results show a weaker association for British journals. Historical precedent, the use of boards of advisory editors that play both editorial and refereeing roles, and other factors certainly weaken the association between the magnitude of P and choice of decision structure.

The third variable of the manuscript evaluation process, the correlation between referees' recommendations, measures agreement in referees' assessments of submitted papers; one might expect it to be a good index of consensus on standards of scholarship (Gordon 1978). There is good reason to believe, however, that this expectation is incorrect. As noted earlier, the correlation between referees' recommendations may be viewed as a reliability coefficient, and reliability coefficients are influenced not only by the amount of error inherent in a measuring procedure, but also by the variation in the cases being measured (Nunnally 1967, pp. 221–22). Thus, the correlation between referees' recommendations will be larger if they evaluate manuscripts of widely varying publishability than if they evaluate a more homogeneous set. It seems plausible that, in fields with little consensus on standards of scholarship, manuscripts of more heterogeneous quality will be submitted for publication than in fields with high consensus. The evidence cited in footnote 14 supports this supposition. If so, the greater disparities among referees' assessments of a given manuscript may be offset by the greater heterogeneity in the publishability of submitted manuscripts, reducing variation across fields in the correlation between referees' recommendations.

Although there is no quantitative evidence on the relative heterogeneity of manuscripts submitted for publication in various fields, Cole and Cole (1981) report such data for grant proposals submitted to three programs in the National Science Foundation. They find greater between-proposal variance in reviewer evaluations in Economics than in Chemical Dynamics or Solid State Physics. Within-proposal variance in reviewers' evaluations is also greater in Econom-

ics, and, as a result, the ratio of between-proposal to total variance in reviewer evaluations (another type of reliability coefficient) has similar values in all three fields. Thus, even though economists disagree more in evaluating a given proposal (the within-proposal variance), the reliabilities of evaluations in the three fields are similar because proposals submitted to the economics program are more heterogeneous.

So the correlation between referees' evaluations is unlikely to be a good measure of consensus on standards of scholarship because it is sensitive to the degree of dispersion in the publishability of submitted manuscripts. This dispersion is probably negatively related both to consensus and the proportion of papers editors summarily reject. When this proportion is high, and when editors correctly anticipate the recommendations they would have received from referees, the dispersion in manuscripts submitted to referees will be greatly reduced, and the correlation between referees' recommendations will be low.

The above considerations suggest that, of the three variables of the manuscript-evaluation process, P is likely to be the most closely linked to consensus. If so, then interdisciplinary variation in journal rejection rates is also probably linked to variation in consensus because variation in rejection rates is primarily a function of variation in P. Indeed, it may be that rejection rates are nearly as good an indicator of consensus as P. Although rejection rates are also affected by the other two variables of the manuscript evaluation process, they are less subject than P to possible distorting effects of the proportion of summary rejections.

The results of this investigation suggest that journal rejection rates cannot be discounted as simply due to journal space shortages and that the greatly different prospects scholars face when submitting manuscripts to, for example, the *Physical Review* and the *American Sociological Review*, result from structural differences between the scholarly communities to which they belong. The effects of interdisciplinary variation in consensus range from day-to-day experiences in doing research (Hargens 1975) to the organizational patterns that typify different fields (Whitley 1984). To neglect the significance of the rejection rate variation is to overlook a fundamental source of the differences between scholarly communities.

Appendix. Average Annual Acceptance Rates and Submissions for 30 Journals During the Late 1960s and Early 1980s

Journal	Average Acceptance Rates		Annual Submissions	
	Late 1960s	Early 1980s	Late 1960s	Early 1980s
Amer. Anthrop.	.31	.15	153	137
Amer. Econ. Rev.	.17	.16	643	748
Amer. J. Phys. Anth.	.75	.78	115	178
Amer. Pol. Sci. Rev.	.17	.11	294	415
Amer. Soc. Rev.	.12	.11	386	475
Analytical Chem.	.59	.67	910	1063
Biochemistry	.73	.65	937	1590
Genetics	.53	.65	280	232
I & E Chem Found.	.55	.56	241	176
I & E Chem Process	.62	.64	165	213
I & E Chem Product	.73	.83	128	163
Inorg. Chem.	.69	.82	948	1122
J. Abnormal Psychol.	.40	.18	298	358
J. Ag. & Food Chem.	.86	.76	310	454
J. Amer. Chem. Soc.	.61	.59	3269	3101
J. Appl. Psychol.	.34	.17	266	567
J. Couns. Clin. Psy.	.25	.15	612	953
J. Edu. Psychol.	.22	.21	351	409
J. Exper. Psychol.	.59	.30	645	646
J. Health & Soc. Beh.	.22	.16	141	189
J. Org. Chem.	.81	.78	1434	1741
J. Pers. & Soc. Psy.	.21	.21	774	1085
J. Phys. Chem.	.65	.76	1425	1291
Macromolecules	.84	.81	190	513
Phys. Rev.	.83	.79	3610	3916
Phys. Rev. Lett.	.48	.44	2141	2414
Psychol. Bull.	.27	.15	265	505
Psychol. Rev.	.26	.14	175	198
Soc. of Edu.	.21	.13	134	139
Soc. Psychol. Quart.	.20	.18	166	154

REFERENCES

Adams, Jack E. 1977. "The Challenge and Response of Economics Journal Politics: A Comparative Survey." *Collegiate News and Views* 30:25–27.

Bakanic, Von. 1986. *Tracing Social Science: The Manuscript Review Process*. Unpublished doc. diss., University of Illinois.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20:37–46.

Cole, Jonathan R. and Stephen Cole. 1981. *Peer Review in the National Science Foundation: Phase Two of a Study*. Washington, DC: National Academy Press.

Cole, Stephen. 1978. "Scientific Reward Systems: A Comparative Analysis." Pp. 167–90 in *Research in Sociology of Knowledge, Sciences and Art* vol. I, edited by Robert A. Jones. Greenwich, CT: JAI Press.
———. 1983. "The Hierarchy of the Sciences?" *American Journal of Sociology* 89:111–39.

Crane, Diana. 1967. "The Gatekeepers of Science: Social Factors Affecting the Selection of Articles for Scientific Journals." *American Sociologist* 2:195–201.

Garvey, William D., Nan Lin, and Carnot E. Nelson. 1970. "Some Comparisons of Communication Activities in the Physical and Social Sciences." Pp. 61–84 in *Communication Among Scientists and Engineers*,

edited by Carnot E. Nelson and Donald K. Pollock. Lexington, MA: D.C. Heath.

Glenn, Norvall D. 1971. "American Sociologists' Evaluations of Sixty-Three Journals." *American Sociologist* 6:298–303.

Gordon, Michael D. 1978. *A Study of the Evaluation of Research Papers by Primary Journals in the U.K.* Leicester, England: Primary Communications Research Center, University of Leicester.

Goudsmit, Samuel A. 1969. "What Happened to My Paper?" *Physics Today* 22:23–5.

Hagstrom, Warren O. 1965. *The Scientific Community*. New York: Basic Books.

Hargens, Lowell L. 1975. *Patterns of Scientific Research: A Comparative Analysis of Research in Three Scientific Fields*. Washington, DC: American Sociological Association.

Hargens, Lowell L. and Warren O. Hagstrom. 1982. "Scientific Consensus and Academic Status Attainment Patterns." *Sociology of Education* 55:183–96.

Hawkins, Robert G., Lawrence S. Ritter, and Ingo Walter. 1973. "What Economists Think of Their Journals." *Journal of Political Economy* 81:1017–32.

Huber, Bettina J. 1982. *Publishing Options: An Author's Guide to Journals*. Washington, DC: American Sociological Association.

Jones, Charles O. 1977. "Report of the Managing Editor of the American Political Science Review." *PS* 10:448–53.
———. 1978. "Report of the Managing Editor of the American Political Science Review." *PS* 11:506–10.

Keyfitz, Nathan. 1975. "How Do We Know the Facts of Demography?" *Population and Development Review* 1:267–88.

King, Donald W., Dennis D. McDonald, and Nancy K. Roderer. 1981. *Scientific Journals in the United States*. Stroudsburg, PA: Hutchinson Press.

Knorr-Cetina, Karin D. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. New York: Pergamon Press.

Merton, Robert K. 1957. "Priorities in Scientific Discovery." *American Sociological Review* 22:635–59.

Mullins, Carolyn J. 1977. *A Guide to Writing and Publishing in the Social and Behavioral Sciences*. New York: Wiley.

Nunnally, Jum C. 1967. *Psychometric Theory*. New York: McGraw-Hill.

Oster, Sharon. 1980. "The Optimal Order for Submitting Manuscripts." *American Economic Review* 70:44–48.

Peters, Douglas P. and Stephen J. Ceci. 1982. "Peer Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again." *Behavioral and Brain Sciences* 5:187–95.

Pfeffer, Jeffrey, Anthony Leong, and Katherine Strehl. 1977. "Paradigm Development and Particularism: Journal Publication in Three Scientific Disciplines." *Social Forces* 55:938–51.

Showalter, English. 1984. "Who Writes for the *PMLA*?" *Publications of the Modern Language Association* 99:147–49.

Smigel, Erwin O. and H. Lawrence Ross. 1970. "Factors in the Editorial Decision." *American Sociologist* 5:19–21.

Stinchcombe, Arthur L. and Richard Ofshe. 1969. "On Journal Editing as a Probabilistic Process." *American Sociologist* 4:116–17.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185:1124–31.

Whitley, Richard D. 1970. "The Formal Communication System of Science: A Study of the Organization of British Social Science Journals." Pp. 163–179 in *The Sociological Review, Monograph 16,* edited by Paul Halmos. Keele, England: University of Keele.

_____. 1984. *The Intellectual and Social Organization of the Sciences.* London: Oxford University Press.

Ziman, John. 1968. *Public Knowledge: The Social Dimensions of Science.* Cambridge: Cambridge University Press.

Zuckerman, Harriet A. and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System." *Minerva* 9:66–100.

DO JOURNAL REJECTION RATES INDEX CONSENSUS?*

(Comment on Hargens, *ASR*, this issue)

STEPHEN COLE
SUNY at Stony Brook

GARY SIMON
New York University

JONATHAN R. COLE
Columbia University

Hargens has made a significant contribution to our continuing attempt to understand the cognitive and social differences between scientific disciplines. A primary issue is to what extent does the level of cognitive consensus in the natural and the social sciences differ? Hargens (1975) and most others (e.g., Kuhn 1970, Lodahl and Gordon 1972, Price 1970, Storer 1967, and Zuckerman and Merton 1971; 1973) have long advocated the accepted assumption that the natural sciences have higher levels of consensus than the social sciences. We have argued that, at the "research frontier," where new knowledge is currently being produced (as opposed to the "core" where a small set of theories and analytic techniques represent the given at the time), all sciences have similar levels of relatively low consensus (S. Cole 1978; S. Cole, J.R. Cole, and Dietrich 1978; S. Cole 1979; S. Cole, J.R. Cole, and G. Simon 1981; and, primarily, S. Cole 1983).

Hargens' article is an important new contribution to this debate. He focuses on journal rejection rates that are generally higher in the social sciences than in the natural sciences. He interprets these data as supporting the hypothesis that large differences exist in cognitive consensus, disconfirming Cole's (1983) theory. To stimulate further research on this topic, we raise several questions about the conclusions Hargens has drawn from his data and model.

Hargens argues that the lower rejection rate in natural sciences journals is evidence that these fields have more cognitive consensus. We disagree: He believes that journal rejection rates are a better indicator of field differences in cognitive consensus than σ_R (defined formally below), the extent to which independent evaluators of a scientific product like a journal article or a research proposal reach the same conclu-

sion. We believe that the latter measure is the best way to compare levels of consensus, and we view journal rejection rates as an inadequate indicator of consensus because they are influenced by consensus *and* other variables. Our comments are divided into two sections: the problems with journal rejection rates as an indicator and the merits of σ_R as an indicator. We also point out two problems in the formal model presented by Hargens.

JOURNAL REJECTION RATES AS A MEASURE OF CONSENSUS

S. Cole (1983) theorizes that the data on journal rejection rates would represent an anomaly if it could be empirically demonstrated that rejection rates were primarily a result of the level of cognitive consensus in a field. Thus far, no one has presented any evidence to support this assumption. Hargens would probably agree that journal rejection rates are influenced by consensus and other variables. Our problem is to measure the relative influence of all the significant variables.

Hargens presents data on another variable, the amount of space available in journals, that might influence rejection rates. We and others have suggested that *one* reason for the higher social-science rejection rate may be less space available in these disciplines. Hargens attempts to disprove this by showing that changes in the number of submissions over time does not explain much variance on acceptance rate. His argument would be more convincing if he had data on the key variable: the number of pages available (adjusted for the average length of articles) divided by the number of people wanting to publish. In 1975, Hargens presented some direct evidence that showed that the ratio of mean circulation to mean number of articles published was more than twice as high in social science journals than in chemistry journals and that the backlogs of unpublished articles were substantially greater in social science journals. In his current paper, he dismisses these data because they are "probably contaminated by other variables, such as levels of competition for priority." The shortage of space variable

deserves a more direct test before being dismissed completely. We agree with Hargens, however, that space shortages alone cannot explain the large differences in rejection rates.

Another important variable influencing rejection rates is field-specific norms concerning publication. Hargens mentions this variable, but he includes it only in so far as the field norms influence the decision structure used by the journals. "Zuckerman and Merton (1971) note that journals with high acceptance rates tend to use procedures that *presuppose* that submitted papers should be published and to minimize the chance that worthy papers will be rejected . . ." (p. 141, emphasis added; see also Bakanic, McPhail, and Simon (1987, p. 633). Some fields, such as physics, have a norm that submitted articles should be published unless they are wrong. They prefer to make "Type I" errors of accepting unimportant work rather than "Type II" errors of rejecting potentially important work. This policy often leads to the publication of trivial articles with little or no theoretical significance, a reason frequently cited by referees in social science fields in rejecting articles. Other fields, such as sociology in the U.S., have a norm that an article should be · rejected unless it represents a significant contribution to knowledge. Sociologists prefer to make Type II errors.

Hargens concludes that "editors rarely report that they have been forced to reject more papers than the reviewing process dictates." This is not surprising if we assume that the reviewers, aware of and influenced by the norms, provide the editors with opportunity to conclude that the articles submitted are not publishable. For example, the editor of *Social Forces* tells the reviewers explicitly that only about 15 percent of articles submitted can be accepted. A reviewer receiving such a letter will look for reason to reject. There are much qualitative data to support the publication-norm hypothesis. For example, the most important sociology journal in Poland, *Studia Sociologiczne,* published by the Polish Academy of Sciences without government interference, accepts a majority of the papers submitted.[1] In Poland, sociologists do not subscribe to the norm that articles should be rejected unless they are a significant contribution. But Hargens would have to interpret these data to mean that there is more cognitive consensus among Polish sociologists than among American sociologists.

The publication-norm hypothesis can be empirically tested. Instead of sending articles to referees and asking them to decide whether they should be published (a decision influenced by the norms), send random samples of submitted articles to qualified reviewers (one paper per reviewer) and ask them to rate the articles on a numeric scale. Then examine the estimate of σ_R, the reviewer standard deviation (discussed below), which measures consensus or the extent to which the raters in various fields agree on the merits of the submitted articles. If such a study is done and there are significant differences in σ_R between the natural and social sciences, Hargens would be right and we would be wrong.

A third variable dismissed by Hargens that might influence rejection rates is the diffuseness of a field's journal system. In physics, the two leading journals publish more than 50 percent of all literature in the field; in sociology, the two leading journals publish only a small fraction of all literature. This means that sociologists must submit their articles to one journal after another until the article is published or the author gives up. Bakanic, McPhail, and Simon (1987) point out that, of all manuscripts submitted to *ASR*, "nearly 29 percent were eventually published" (p. 640). As Garvey, Lin, and Nelson (1970) have pointed out, more than 20 percent of articles published in "core" social science journals had previously been rejected by one or more journals. Thus, some articles rejected by *ASR* will be published by the *American Journal of Sociology, Social Forces,* or some other core journal. Although we do not have any data on this, it is quite likely that at least half the articles submitted to *ASR* are eventually published in a core journal and that probably three quarters of them are eventually published somewhere. If *ASR, AJS, SF,* and all the other journals published by the ASA were, in fact, one journal, a situation similar to that in physics, it is probable that the rejection rate in sociology would not differ significantly from that in some of the natural sciences.

A fourth variable that might explain differences in rejection rates could be differences in training practices in various fields. In the natural sciences, students tend to work more under the direction of a sponsor than in the social sciences. Students generally work on problems selected by their sponsors and write their first articles in collaboration with their sponsors. Dissertations are more likely to be articles than books, and it is probable that students learn to write a publishable article by collaborating with their mentors. Social science disciplines are less efficient in teaching students how to write publishable articles. Students are more likely to select their own problems, work independently of their sponsors, and have relatively little

---

[1] The Polish journal does not use blind refereeing (neither do U.S physics journals). The articles are refereed by members of the editorial advisory board. Also, like U.S. physics journals, the Polish journal probably does not have as much competition for space as do U.S. sociology journals.

guidance when preparing their first articles. As a result of these differences in training, scientists in the social sciences may be more likely to submit inadequate papers than scientists in the natural sciences.

Hargens treats these variables by arguing that they are intervening between consensus and rejection rates, i.e., they are caused by differences in consensus. For example, he implies that the fact that *ASR* and *AJS* publish only a small portion of the literature in sociology and *PR* and *Physical Review Letters* publish a large portion of the literature in physics is a result of differences in consensus between the two fields. However, until there is some *direct* measure of consensus, this argument simply restates the *assumption* that the natural sciences have more consensus than the social sciences. Direct measures of consensus in these fields do exist.

MEASURING CONSENSUS

One problem in Hargens' article is his failure to specify adequately the meaning of *consensus* and how it should be measured. Hargens' model uses three variables to predict acceptance rates: the type of decision structure used by the editor, the level of agreement between reviewers of the same article, and the total proportion of reviews recommending acceptance. Although he points to some interesting interactions, he concludes that the last variable is the most significant. The finding that those journals in which a high proportion of reviewers recommend acceptance have a high proportion of articles accepted is not particularly informative. Given a low P, the proportion of reviews recommending acceptance, a journal *must* have a *relatively* high rejection rate regardless of the other variables in the model (see Hargens' Table 3). This does not tell us why P differs from field to field. Hargens offers some interesting speculation on this question, but no empirical evidence. P could be primarily a result of level of cognitive consensus as he argues, or it could also be influenced by any of the four variables we discussed above or other variables.

We define consensus as the extent to which scientists in a field agree on intellectual issues. It should be measured by the ability of independent referees to reach similar conclusions about the value of journal articles, research proposals, or other scientific products. Instead of using the extent of reviewer agreement as a direct measure of consensus, Hargens prefers the rejection rate, which is influenced by both extent of reviewer agreement and many other variables. Let us examine why Hargens does not prefer extent of reviewer agreement as a measure of consensus.

Hargens' points out that most statistical measures used in the past to assess the extent to which referees agree are inadequate *comparative* measures of consensus because they are influenced by the heterogeneity of items being evaluated (papers or research proposals) as well as by the consensus of reviewers on the merits of each item. Hargens refers to our analysis of consensus among referees of proposals submitted to the National Science Foundation (Cole and Cole 1981, Cole, Cole, and Simon 1981). Here we compared the proportion of variance resulting from referee disagreement on given proposals with the proportion of variance resulting from differences among the proposals (as assessed by the mean rating each proposal received.) In S. Cole (1983), roughly equal ratios for economics, biochemistry, and solid-state physics support the hypothesis that the levels of cognitive consensus in the natural and social sciences are roughly equal. Hargens argues that this ratio is influenced by the greater heterogeneity of proposals and articles in the social sciences and that lower levels of consensus in these fields will thus yield equivalent ratios. We agree with the statistical point, but data presented in the second column of Table 6 (S. Cole 1983) show the average standard deviation of ratings within proposals, a statistic that is not a ratio and therefore not influenced by field differences in proposal heterogeneity. This statistic is a valid measure of the extent to which scientists agree and is appropriate for making cross-field comparisons. These data show that at the individual-proposal level, economics has more cognitive consensus than some of the natural sciences and that anthropology has cognitive consensus at the same level as biochemistry and geophysics.

In our opinion, the parameter that connotes consensus is σ_R, the standard deviation associated with the reviewer noise. In Cole, Rubin, and Cole (1978) and Cole and Cole (1981), the estimated values of this parameter were found to vary only modestly over a range of social science and natural science disciplines. Because the between-proposal standard deviation σ_P is influenced by a large number of variables other than consensus, it follows that all calculations that use this quantity σ_P are not useful in making comparative analyses of the level of consensus in various scientific fields. These calculations include the F-statistic, the R^2 statistic, the between-to-within variance ratio, the intraclass correlation and so on. Of course, these calculations do indicate the ability of the refereeing process to discriminate publishable and nonpublishable papers (or fundable and nonfundable research proposals), given the character of papers that are submitted. (For example, large

R^2 values indicate easier accept/reject decisions.)

Hargens' argument that varying levels of heterogeneity among the population of submitted articles influence the ease of reaching consensus on an individual article does not seem relevant to the way in which journal articles or NSF proposals are evaluated. His point would be more convincing if a group of referees were evaluating a *group* of articles. In that circumstance, large heterogeneity between the articles would make it easier to reach a consensus in ranking. But when several independent reviewers evaluate a *single* article, it is not evident how heterogeneity can influence the difficulty of evaluating the single article or the extent to which independent reviewers will reach a similar conclusion. Further, only slim evidence supports the contention that heterogeneity of submissions in the social sciences is greater than in the natural sciences. Our data on peer review do show *slightly* more heterogeneity of submitted NSF proposals for the social sciences. Hargens assumes that there is more heterogeneity among the 85 percent of articles rejected by ASR than among the 85 percent of articles accepted by PR. The only evidence he cites is Zuckerman and Merton's (1971) report that editors of high rejection-rate journals are more likely to reject papers without outside review.

Because Hargens (p. 146) dismisses reviewer agreement as a measure of consensus since it is influenced by heterogeneity, he is forced to argue that P is a result of consensus between referees and authors. All authors probably believe that their articles deserve to be published, so that P becomes solely a function of the proportion of reviewers giving negative reviews.

We believe that the most compelling evidence against Hargens' argument may be found in our peer review study (S. Cole, J. R. Cole, and Simon 1981). Here we report that three fields had very similar levels of σ_R: about 7 in solid-state physics, about 7.5 in chemical dynamics, and about 9 in economics. Our problem in the sociology of science has been and continues to be that we have incorrectly *assumed* that there were high levels of consensus in the natural sciences. Since there are low levels of consensus in the social sciences, we have assumed field differences. Now that we have good direct data on the level of cognitive consensus in the natural sciences, we question the validity of this assumption.

Hargens' data support our contention that there are not meaningful differences in consensus in evaluating journal articles in physics and in sociology. Consider Hargens' data on articles submitted to ASR. If disagreement in the field over what constitutes a publishable article is the

primary cause of the high rejection rates, we should see evidence of that disagreement in many split decisions. In fact, in almost 60 percent of the cases (see Hargens' Table 4B), both referees agree that the article is not publishable (and in 10 percent that it is publishable). Hargens points out (note 12) that there is a small positive correlation between the opinions of independent ASR referees.[2] He also points out that the data on PR are compatible with the interpretation that the opinions of referees (in those cases where more than one was used) are statistically independent. Both these data and our peer review data lead to the conclusion that there are *not* meaningful differences in the extent of agreement among independent evaluators in fields like physics and sociology.

Finally, we believe there are two problems in Hargens' model. First, the assertion that a single P is associated with each journal seems logically unsupportable. Consider the data of Table 4B, ASR Referees. This shows the breakdown:

Both say accept	34	(10%)
Split	114	(33%)
Both say reject	194	(57%)
TOTAL	342	

If the author's model is correct and if the referees operate independently, then the proportions should be in the ratio $P^2 : 2PQ : Q^2$, where $Q = 1 - P$. The maximum likelihood estimate for $P =$ probability of a favorable review is (2 $\times$ 34 + 114) / (2 $\times$ 342) = 0.2661. This would lead to the table:

	Observed	Fitted
Both say accept	34	24.21
Split	114	133.57
Both say reject	194	184.21
TOTAL	342	341.99

The resulting chi-squared statistic is 7.34 on one degree of freedom. (There is one d.f. because the value of P was estimated from the data.) Since the chi-square exceeds 6.63, the model (with a single P and "statistically independent" judges) must be rejected at the .01 level. The data have too few split decisions. This could indicate that there are different Ps for different papers (which seems intuitively correct to us),

---

[2] In note 12, Hargens uses Cohen's kappa to measure the extent of the agreement between referees. Cohen's kappa is inappropriate for disentangling consensus from proposal heterogeneity because it is a disguised intraclass correlation coefficient (see Fleiss and Cohen 1973) and thus reflects between-paper as well as within-paper variability. Maclure and Willett (1987) note that "A characteristic of kappa . . . is its variation with changes in prevalences of the phenomenon being measured" (p. 165). Specifically, a heterogeneous set of papers will tend to produce a large kappa.

that the referees' judgments are not "statistically independent" (an implausible assumption unless the referees collude), or even that there is some bias introduced by the editorial process.

Second, from a modeling point of view, a two-stage process seems to be at work. First, some random mechanism creates papers that are sent for review and then the papers are sent to reviewers for independent appraisals. If a reviewer's work produces a numerical (rather than a yes/no) score, then we should think of

$$S_{ij} = m + p_i + r_{ij}$$

where S_{ij} is the score given to paper i by its jth reviewer, m is an overall mean, p_i is the random effect due to paper i and r_{ij} is the noise contributed by the reviewer. The most interesting parameter is σ_R, the standard deviation of the reviewer noise, which measures the consensus in the field. Also interesting is the standard deviation of the proposal merit, which measures the heterogeneity of the proposals submitted for review. Note that if σ_P, the standard deviation associated with P_i, is large relative to σ_R, then it is easy for the editorial process to make publication decisions. Statistical calculations that use ratios of estimates of σ_R and σ_P quantify the ability to make publication decisions. We maintain that easy decision making is a consequence of both σ_R and σ_P and does not speak to the issue of consensus per se.

If the review of a paper is to result in a simple reject/accept decision, then the S_{ij} model cannot be used. There is a mechanism that generates a value of p, independently for each paper. Once the value of p is generated, the reviewers will independently appraise the paper and find it favorable with probability p. The mechanism that generates p can be modelled as a member of the beta family of probability densities; the two parameters of the density can be used to summarize the journal's papers.

REFERENCES

Bakanic, Von, Clark McPhail, and Rita J. Simon. 1987. "The Manuscript Review and Decision Making Process." *American Sociological Review* 52:631–42.

Cole, Jonathan R. and Stephen Cole. 1981. *Peer Review in the National Science Foundation: Phase Two of a Study*. Washington, DC: National Academy Press.

Cole, Stephen. 1978. "Scientific Reward Systems: A Comparative analysis." Pp. 167–90 in *Research in the Sociology of Knowledge, Science and Art*, edited by R. A. Jones. Greenwich, CT: JAI.

Cole, Stephen. 1979. "Age and Scientific Performance." *American Journal of Sociology* 84:958–77.

Cole, Stephen. 1983. "The Hierarchy of the Sciences?" *American Journal of Sociology* 89, 111–39.

Cole, Stephen, Jonathan R. Cole, and Lorraine Dietrich. 1978. "Measuring the Cognitive State of Scientific Disciplines." Pp. 209–51 in *Toward a Metric of Science: The Advent of Science Indicators*, edited by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, and H. Zuckerman. New York: Wiley.

Cole, Stephen, Jonathan R. Cole, and Gary Simon. 1981. "Chance and Consensus in Peer Review." *Science* 214, 881–86.

Cole, Stephen, Leonard Rubin and Jonathan R. Cole. 1978. *Peer Review in the National Science Foundation: Phase One of a Study*. Washington, DC: National Academy Press.

Fleiss, J.L. and J. Cohen. 1973. "The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability." *Educational Psychology Measurement* 33, 613–19.

Garvey, William D., Nan Lin, and Carnot E. Nelson. 1970. "Some Comparisons of Communication Activities in the Physical and Social Sciences." Pp. 61–84 in *Communication Among Scientists and Engineers*, edited by Carnot E. Nelson and Donald K. Pollock. Lexington, MA: Heath.

Hargens, Lowell. 1975. *Patterns of Scientific Research: A Comparative Analysis of Research in Three Fields*. Washington, DC: American Sociological Association.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press.

Lodahl, Janice B. and Gerald Gordon. 1972. "The Structure of Scientific Fields and the Functioning of University Graduate Departments." *American Sociological Review* 37:57–72.

Maclure, Malcolm and Walter C. Willett. 1987. "Misinterpretation and Misuse of the Kappa Statistics." *American Journal of Epidemiology* 126, 161–69.

Price, Derek J. de Solla. 1970. "Citation Measures of Hard Science, Soft Science, Technology, and Non-Science." Pp. 1–12 in *Communications Among Scientists and Engineers*, edited by Carnot E. Nelson and Donald K. Pollack. Lexington, MA: Heath.

Storer, Norman. 1967. "The Hard Sciences and the Soft: Some Sociological Observations." *Bulletin of the Medical Library Association* 55:75–84.

Zuckerman, Harriet and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System." *Minerva* 9:66–100.

Zuckerman, Harriet and Robert K. Merton. 1973. "Age, Aging and Age Structure in Science." Pp. 497–559 in *The Sociology of Science* by R. K. Merton, edited by Norman Storer. Chicago: University of Chicago Press.

FURTHER EVIDENCE ON FIELD DIFFERENCES IN CONSENSUS FROM THE NSF PEER REVIEW STUDIES*

LOWELL L. HARGENS

University of Illinois at Urbana-Champaign

Cole, Simon, and Cole (CSC) raise a number of objections to my analysis, some pertaining to points I make in my paper and others pertaining to the inconsistency between my conclusions and conclusions they have drawn elsewhere. I begin by discussing their objections to aspects of the paper itself and then turn to the inconsistency between some of my conclusions and theirs.

ALTERNATIVE EXPLANATIONS OF JOURNAL REJECTION RATES

In my paper I identify three proximate determinants of journal rejection rates and suggest how each may be related to disciplinary variation in consensus. CSC argue that I either ignore or improperly discount variables other than consensus that may be responsible for the observed interdisciplinary variation in journal rejection rates.

For example, CSC argue that my analysis of the impact of journal space shortages on rejection rates is unconvincing because it does not measure journal-space availability directly. I believe they greatly underestimate the difficulties involved in obtaining such a measure— difficulties that frequently occur in attempts to measure potentials (rather than actualities).[1] but even if an adequate measure were available, simply comparing its average value across disciplines would be uninformative because the issue is one of causal priority. Specifically, the space-shortage argument holds that both submissions to a journal and its available space jointly produce its rejection rate. If this argument is

correct, a journal's rejection rate should display considerable instability because annual fluctuations in submissions are unlikely to be consistent with the amount of space its annual budget makes available. In contrast, I argue that, for most scientific journals, the volume of submissions and the rejection rate jointly produce the number of pages that are eventually made available for accepted papers. Under this argument one should observe instability primarily in the number of papers a journal publishes because fluctuations in submissions and a fairly constant rejection rate will produce such variation. The data summarized in Table 1 of my paper show that journal rejection rates are extremely stable, and I therefore conclude that the space-shortage argument cannot account for much of the variation in rejection rates. Readers should bear in mind that at issue here is the *interdisciplinary* variation in average journal rejection rates. Until now, Coles et al. have discounted the argument that this variation suggests interdisciplinary variation in consensus, arguing instead that it may be due to variation in space shortages (cf. Zuckerman and Merton 1971, note 35; Cole, Cole, and Dietrich 1978; Cole 1983). However, if space shortages contributed appreciably to the interdisciplinary variation Zuckerman and Merton reported, it is difficult to believe that scholarly associations that publish journals with high rejection rates would not have moved in the intervening years to ameliorate the situation.

CSC identify three additional variables that may affect rejection rates—field-specific publication norms,[2] diffuseness of journal structures, and differences in training practices—and argue that I inappropriately view them as intervening variables that mediate the causal impact of consensus on rejection rates. Specifically, they argue that, in the absence of a direct measure of

---

* Direct all correspondence to Lowell L. Hargens, Department of Sociology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[1] I discussed some of these issues at greater length in Hargens (1975, pp. 20–22) and concluded there that the existing measures are all indirect and allow only the conclusion that space shortages are not the *sole* source of rejection rates. Although CSC now agree with my earlier conclusion, I believe that the evidence I present in the current paper justifies the much stronger conclusion that space shortages have little to do with the observed variation in rejection rates.

[2] Although CSC argue for the existence of *field*-specific publication norms, their examples, *Social Forces* and *Studia Sociologiczne*, imply that *journal*-specific norms are involved. As far as the latter journal is concerned, I doubt whether it is appropriate to compare it with U.S. sociology journals and suspect that a more appropriate comparison would be with Polish physics journals.

consensus in different fields, the claim that these three variables intervene between consensus and rejection rates is "nothing more than a restate-ment of the *assumption* that the natural sciences have more consensus than the social sciences" [emphasis in the original]. Aside from noting that this seems to be another instance of the fallacy of believing that one can specify correct causal orderings on the basis of observed associations, I must point out that the ordering I posit is not an arbitrary and idiosyncratic one. For example, Kuhn (1970) and Lodahl and Gordon (1972) discussed the effect of consensus on differences in training patterns, and Zucker-man and Merton (1971) suggested that variation in the decision rules used in evaluating manu-scripts may be a function of variation in consensus. CSC reject these arguments and findings because they are not based on direct measures of consensus, but CSC do not explain why the three variables they identify covary across disciplines. My claim is that interdiscipli-nary variation in consensus can account for the covariation, and rejecting this claim just because investigators have used only indirect measures of consensus seems surprisingly positivistic.

PROBLEMS IN THE FORMAL MODEL

CSC argue that my use of a single P to characterize a given journal is "logically unsupportable" because the papers submitted to any journal are heterogeneous in their publish-ability. I agree that such heterogeneity exists, and that the chi-square test I report in note 12 of my paper suggests its presence among papers submitted to *ASR*. However, variation in *papers'* publishability, p, does not logically preclude characterizing a *journal* in terms of its proportion of positive referee recommendations, P. I used the latter to determine how structural features of the reviewing process affect various outcomes of that process. For that purpose, whether papers are of heterogeneous publish-ability is irrelevant. Thus, the criticisms CSC make on this point seem to stem from their misunderstanding the point of my model.

The modeling approach CSC recommend in the final paragraphs of their comment is one of many possible alternatives for studying the sources of variation in P, but it immediately raises the thorny issue of the meaning and sources of variation in p. The original question that motivated my paper was to explain why disciplines vary so widely in average journal rejection rates. To conclude that this variation is caused by variation in average paper publish-ability levels does not take us very far toward answering the question.

INCONSISTENCIES BETWEEN MY CONCLUSIONS AND THOSE PREVIOUSLY DRAWN BY CSC

The main objections CSC raise against my analysis stem from my suggestion in the first part of the final section of my paper that disciplinary variation in average levels of P are a function of disciplinary variation in consensus. Specifically, they contend that my findings and arguments should be dismissed because their reports on studies of peer review at the National Science Foundation show results for a direct measure of consensus that varies little across disciplines. I believe that the NSF peer-review results are much more equivocal than CSC suggest and that they do not justify rejecting my suggestions about possible relations between consensus and P.

Let us begin with the issue of whether the Coles' peer-review studies yielded direct mea-sures of consensus. In part, this is a conceptual issue. The measure CSC now espouse, σ_R, is based on a conception of consensus that focuses only on individual agreement rather than taking higher levels of co-orientation into account. Scheff (1967) effectively criticized the individual-agreement model and argued for an interaction-ist model that incorporates higher levels of individuals' understandings about the existence or absence of agreement. I followed Scheff's approach in my previous discussions of the concept of consensus (cf. Hargens 1975), which I probably should have repeated in this paper. In any case, I believe that measures of consensus based on agreement alone, such as σ_R, are indirect measures. Moreover, CSC are in error when they claim that I believe journal rejection rates, but not σ_R, to be a proper measure of consensus. Both kinds of measures, though indirect, are proper, and both should reflect variation in consensus.

CSC argue that the most compelling evidence against the claim that fields vary in consensus is found in the data from Phase Two of their analysis of the evaluation of grant proposals submitted to the NSF (Cole and Cole 1981). In that study, they obtained measures of σ_R from three data sets for each of three fields: economics, chemical dynamics, and solid-state physics. For the first data set (the original NSF referees), the values of σ_R are approximately 9.5 (not 9.0 as they state in their comment) for economics, 7.5 for chemical dynamics, and 7.0 for solid-state physics. Although these differ-ences are consistent with economics having less consensus than the two physical sciences, CSC conclude that the three fields have "very similar" levels of σ_R. Statistical tests do not sustain their conclusion. For example, a simple test can be calculated using the reported values

of σ_R from the three different data sets (Cole and Cole 1981, pp. 38–40). For all three data sets, economics shows higher values of σ_R than the other fields, and the probability of this happening under the null hypothesis that the three fields have equal levels of consensus equals $(1/3)^3$, or .037. Taking into account the *magnitudes* of the field differences in within-proposal rater variances yields much higher levels of statistical significance.[3] Thus, statistical decision rules lead one to conclude that, in the Phase Two NSF peer-review data, economics exhibits less consensus than the physical-science fields. Those data can hardly constitute evidence *against* the argument that the social sciences have less consensus than the natural sciences.

Data in column 2 of Table 6 in Cole (1983) provide apparently stronger evidence against the argument that the social sciences have less consensus than the natural sciences. These are data on the variability of referees' ratings of NSF proposals in the 10 fields the Coles examined in Phase One of their study, and they do not show uniformly higher disagreement among referees in the social science NSF programs. Unfortunately, rather than reporting σ_R, Cole reports the "mean standard deviation" of referees' ratings of each of the 100 proposals examined in each of the 10 NSF programs. Because not all NSF proposals are rated by the same number of referees (the range is from one to eight), this measure has two weaknesses. First, it does not equally weight all referees' ratings; instead it gives higher weights to those who rated proposals with few referees. Second, this measure is heavily influenced by proposals that receive only one referee rating because such proposals were assigned a standard deviation

value of zero (see Cole and Cole 1981, p. 74). Ceteris paribus, programs that have larger proportions of proposals with a single referee should show lower values of the mean standard deviation. Cole (1983) reported that the values of this measure for economics, solid-state physics, and chemical dynamics equal .34, .35, and .42 respectively. Although these results appear to be inconsistent with those in the Phase Two report, I discovered on examining the raw data in the latter that 16 percent of the 50 proposals in the economics program received only one referee evaluation while none in the other two programs did (Cole and Cole 1981, Tables A1, A4, and A7).

The NSF peer-review studies are not the only attempt to measure consensus in different fields using the individual-agreement model of consensus. Gwendolyn Lewis (1980) studied academic scientists' agreement on the importance of recent research findings and future directions of research in genetics, solid state physics, and developmental psychology. She found that, for both variables, developmental psychology displayed statistically significantly less consensus than the other two fields.

Thus, two of the three studies of consensus using the individual agreement model of consensus yielded statistically significant results consistent with the argument that the social sciences have less consensus than the natural sciences, and the third (Cole 1983) relied on a measure that may have a serious technical flaw. I believe that these results, although clearly not definitive, provide evidence that is at least weakly favorable to that argument.

Finally, CSC argue that in my paper I report data inconsistent with the argument that the social sciences exhibit less consensus than the natural sciences. Specifically, they interpret data in Table 4 of my paper as suggesting that referee evaluations of papers submitted to *Physical Review (PR)* are statistically independent. Because I show in Table 4 that referee evaluations for *American Sociological Review* are nearly statistically independent, CSC conclude that the data suggest that physics and sociology have similar levels of consensus. However, CSC apparently misread these data because those reported for *PR* in Table 4 are for initial editorial dispositions, not referee recommendations. The only evidence on the association between referee recommendations for *PR* that I know of (Zuckerman and Merton 1971, note 3) indicates a substantial association.

In sum, I believe the evidence CSC cite does not support rejecting my analysis of the relationship between consensus and *P*. In part, our disagreement stems from a problem common in the social sciences, the lack of definitive measures of important concepts. Lacking strong

---

[3] The F-test of equality of variances is very sensitive to departures from the normality assumption (Box, 1953), but alternatives that do not suffer from this defect are available. One alternative Box suggested is to (1) calculate the variance of the referee ratings for each of the proposals in a given field (only proposals with at least two ratings can be included); (2) take the natural logarithms of each of those variances; (3) calculate the mean value of the logged variances for each field and the standard error of those means; and (4) carry out a t-test for the difference between any two means. In the case at hand, I pooled the mean logged variances and their standard errors for chemical dynamics and solid-state physics, and calculated a t-test for the difference between economics and the pooled fields. For those proposals assigned identical scores by all referees, I changed one referee's rating by one-half of the smallest possible difference between scores to avoid taking the logarithm of zero. The t-value yielded by the data for the NSF referees reported by Cole and Cole (1981, Tables A1, A4, and A7) equals 15.07. I thank Lincoln E. Moses for helpful discussions and advice on these issues.

measures, we typically rely on multiple imperfect ones, and this is likely to lead to disputes about the adequacy of evidence adduced for particular substantive claims. This exchange is a case in point. Perhaps a future study should examine the probability that a published paper will provoke a critical comment as a possible measure of scholarly consensus.

REFERENCES

Box, G.E.P. 1953. "Non Normality and Tests on Variances." *Biometrica* 40:318–55.
Cole, Jonathan R. And Stephen Cole. 1981. *Peer Review in the National Science Foundation: Phase Two of a Study*. Washington, DC: National Academy Press.
Cole, Stephen. 1983. "The Hierarchy of the Sciences?" *American Journal of Sociology* 89:111–39.
Cole, Stephen, Jonathan R. Cole, and Lorraine Dietrich. 1978. "Measuring the Cognitive State of Scientific Disciplines" Pp. 209–51 in *Toward a Metric of Science: The Advent of Science Indicators*, edited by Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, and H. Zuckerman. New York: Wiley.
Hargens, Lowell L. 1975. *Patterns of Scientific Research: A Comparative Analysis of Research in Three Scientific Fields*. Washington, DC: American Sociological Association.
Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*, 2nd Ed. Chicago: University of Chicago Press.
Lewis, Gwendolyn L. 1980. "The Relationship of Conceptual Development and Consensus: An Exploratory Analysis of Three Subfields." *Social Studies of Science* 10:285–307.
Lodahl, Janice B. And Gerald Gordon. 1972. "The Structure of Scientific Fields and the Functioning of University Graduate Departments." *American Sociological Review* 37:57–72.
Scheff, Thomas J. 1967. "Toward a Sociological Model of Consensus." *American Sociological Review* 32:32–46.
Zuckerman, Harriet and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System." *Minerva* 9: 66–100.

ASA Journals

◆

Current and Forthcoming Articles

Journal of Health and Social Behavior
Vol. 28, No. 2 (June 1987)

Holistic Physicians: Implications for the Study of the Medical Profession, *Michael S. Goldstein, Dennis T. Jaffe, Carol Sutherland, and Josie Wilson*

Concern with Appearance, Health Beliefs, and Eating Habits, *Diane Hayes and Catherine E. Ross*

Some Factors Associated with the Report and Evaluation of Back Pain, *David Mechanic and Ronald J. Angel*

The Sociological Study of AIDS: A Critical Review of the Literature and Suggested Research Agenda, *Howard B. Kaplan, Robert J. Johnson, Carol A. Bailey, and William Simon*

Gender Role Orientations and Styles of Pathology Among Adolescents, *Allan V. Horwitz and Helene Raskin White*

Gender Differences in the Use of Mental Health-Related Services: A Reexamination, *Philip J. Leaf and Martha Livingston Bruce*

The Influence of Family Caregivers on Elder's Use of In-Home Services: An Expanded Conceptual Framework, *David M. Bass and Linda S. Noelker*

Transitory Impact of Life-Event Stress on Psychological Symptoms in Older Adults, *Fran H. Norris and Stanley A. Murrell*

Journal of Health and Social Behavior
Vol. 28, No. 3 (September 1987)

Prognostic Conflict in Life-and-Death Decisions: The Organization as an Ecology of Knowledge, *Renee R. Anspach*

Social Psychological Aspects of Reporting Behavior: Lifetime Depressive Episode Reports, *Carol S. Aneshensel, Antonio L. Estrada, Mary Jo Hansell, and Virginia A. Clark*

The Work-Family Role System and Physician Productivity, *Carol S. Weisman and Martha A. Teitelbaum*

Change in Student Physicians' Views on Authority Relations with Patients, *Bebe*

Lavin, Marie Haug, Linda Liska Belgrave, and Naomi Breslau

A Multivariate Model of the Determinants of Cigarette Smoking Among Adolescents, *Felipe G. Castro, Ebrahim Maddahian, Michael D. Newcomb, and P.M. Bentler*

Sex Differences in Distress: The Impact of Gender and Work Roles, *Mary Clare Lennon*

Family Status and Health Behaviors: Social Control as a Dimension of Social Integration, *Debra Umberson*

Personal Efficacy, Income, and Family Transitions: A Longitudinal Study of Women Heading Households, *Geraldine Downey and Phyllis Moen*

Teaching Sociology
Vol. 15, No. 4 (October 1987)

Are They Getting What They Expect? *Richard J. Gigliotti*

Building the Commitment to Undergraduate Education: A Structural Response, *Michael Brooks*

Critical Thinking in the Sociology Classroom: Facilitating Movement from Vague Objective to Explicit Achievement, *M. Neil Browne and James L. Litwin*

The Use of Conceptual Analysis for Teaching Sociology Courses, *Marcel Fredericks and Steven I. Miller*

Using Qualitative Data from Children to Teach about Gender Role Socialization, *Joy B. Reeves*

Introducing Sociology at a Distance: Courses for Adults with Minimal Educational Background, *Yael Enoch*

Reducing Classroom Alienation: Applications from Theory, *Barbara R. Keating*

Using Retrospective Autobiographies as a Teaching Tool, *Kim M. King*

The Social Construction of Grades, *Fred P. Pestello*

Are "Sisters in Crime" Finally Being Booked? The Coverage of Women and Crime in Journals and Textbooks, *Richard A. Wright*

Don't Miss This Valuable Resource!

Cumulative Index
of Sociology Journals

The Cumulative Index of Sociology Journals, published in early 1987, covers the 1971–1985 period. Compiled by Judith Lantz of Carbondale, it contains both an author and keyword or subject index. Books reviewed in any of the journals that are part of the Index are included in both the author and subject listings. The following ASA-sponsored publications are included in the 780-page Cumulative Index: American Sociological Review, The American Sociologist, Contemporary Sociology, Journal of Health and Social Behavior, Social Psychology Quarterly, Sociological Methodology, Sociological Theory, and Sociology of Education.

In addition, both the American Journal of Sociology and Social Forces have been incorporated into the index. Their inclusion represents the Association's first, and very preliminary, attempt to put together a single index of all major sociology journals. Future editions will include a larger number of journals.

The Cumulative Index is available to ASA members for $37.50, to non-members for $48.50 and to institutions for $65.00. Send prepaid orders to: ASA Executive Office, 1722 N Street, N.W., Washington, D.C. 20036. Act now to get this valuable new resource!

I

Social Interaction
READINGS IN SOCIOLOGY
Third Edition

CANDACE CLARK, *Montclair State College*
HOWARD ROBBOY, *Trenton State College*

This anthology of 38 readings, designed to sup-
plement virtually any introductory sociology text,
is organized around the symbolic interactionist
perspective. The articles, both classic and con-
temporary, cover such topics as the feminization
of poverty, religion, aging, and career mobility;
17 of the articles are new to this edition. Also new
to this edition are a subsection on the sociology
of health and illness and material on the sociol-
ogy of emotions. Each major section of the book
begins with an introduction that explains how the
following articles fit into the larger sociological
picture. Review questions are provided for each
article.
Paperbound.   608 pages.   *Just Published*
Instructor's Manual available
Instructor's Edition—text and *Instructor's Manual*
bound together in one volume—available

Society and Technological Change
RUDI VOLTI, *Pitzer College*

This narrative text focuses on social, political, and
economic issues throughout the world that affect,
and are affected by, technological change—
including the political dilemma of arms control,
the social controversy regarding violence on tele-
vision, and the economic impact of technology
on the job market. The author explains theories
and specific technologies in plain language and
supports his explanations with case studies,
examples, and illustrations. Each chapter ends
with four or five questions for discussion.
Paperbound.   352 pages (probable).
January 1988
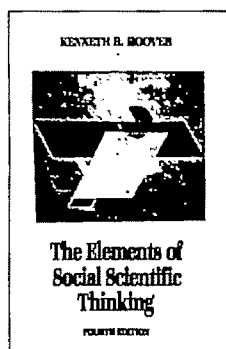
The Elements of Social Scientific Thinking

Fourth Edition

KENNETH R. HOOVER, *University of Wisconsin, Parkside*

A thorough introduction to the vocabulary, concepts and objectives of social scientific thought, this brief text is suitable for a wide range of courses in any one of the social science disciplines. Each chapter begins with an outline of topics to be covered and ends with a review of the major concepts. The appendixes include a model research paper and a demonstration of the application of regression analysis. New to this edition are a fuller explanation of various classes of independent variables, a revised presentation of path analysis, and an expanded discussion of "Morality and the Limits of Science."

Paperbound.   179 pages.   *Just Published*
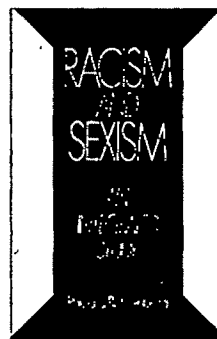
Racism and Sexism

AN INTEGRATED STUDY

PAULA S. ROTHENBERG, *William Paterson College*

This text-reader thoroughly integrates the study of racism and sexism within the context of class in American society. Drawing on theoretical articles from sociology, psychology, history, philosophy, and literature as well as selections from popular sources such as *Ms.* and *Essence*, Professor Rothenberg provides evidence of racism and sexism, covers their impact on society, and offers prospects and strategies for change. Each section begins with an introduction that describes and clarifies the essays to follow.

Paperbound.   404 pages.   *Just Published*

To request an examination copy of any of these titles, please call our toll-free number, 800-221-7945 and ask for Ext. 554; or you may write us on your college letterhead specifying your course title, present text, and approximate enrollment. Send your request to:

St. Martin's Press

III

Presenting the exciting new *Fifth Edition* of a classic sociology text—

DOWN TO EARTH
SOCIOLOGY

FIFTH EDITION—1988

JAMES M. HENSLIN, Southern Illinois University, Edwardsville

Now in its Fifth Edition, this widely used text has been thoroughly revised and updated with nearly *forty percent* important new material, including fourteen new articles that examine ethnic differences in expressing pain, gender differences in the work world, the world of computers, friendship, the homeless on skid row, hockey violence, outlaw motorcyclists, after dark work and activity, garage sales, and the experience of extreme isolation. Compatible with all introductory texts, this new edition of *Down To Earth Sociology* retains the basic structure and appealing mixture of classic and contemporary research which have made past editions so popular. Accessible and provocative, Henslin's book focuses on *participant observation*, which offers students a sense of "being there" as sociologists make their investigations. Each article and section is preceded by a concise introduction, which highlights for both professors and students the important understandings to be gained. Also new to this edition is an Instructor's Manual, written by Henslin, that offers practical teaching suggestions for each article.

FEATURES

- Structurally compatible with all introductory texts
- Does not sacrifice accuracy for comprehension—the selections appeal to student interest while clearly presenting basic sociological concepts
- Provides excellent coverage of social inequality, deviance, and social institutions
- Contains fascinating analyses of varied subcultures
- Each article is edited for uniform size
- Includes correlation charts, name and subject indexes, and a glossary
- An extensive Instructor's Test Bank is included as part of the Instructor's Manual

---

Available Now!   *1988   450 Pages   ISBN: 0-02-914451-5 $13.95   paper*
Instructor's Manual ISBN: 0-02-914452-3

College Professors: *To receive your examination copy, contact your Macmillan representative, or write to:*

Fp THE FREE PRESS
A Division of Macmillan, Inc.
866 Third Avenue • New York, NY 10022 • Attn: Dino Battista

V

Our name has changed,
but our commitment
to the field hasn't.

SOCIOLOGY: An Introduction, *Third Edition*
J. Ross Eshleman
Barbara G. Cashion
Laurence A. Basirico

Highly acclaimed for its balanced, integrated presentation of sociological theory and applications, this popular introductory text has been thoroughly updated, revised, and expanded. The new Third Edition integrates an applied sociological perspective into each chapter to show students the relationship between sociological theory and practice beyond the career level. Accessible to students, it offers a direct, jargon-free style and relevant examples. The outstanding pedagogy includes "Sociology at Work" boxes, "Readings in Applied Sociology" in each Part, an updated appendix on "Exploring a Career in Sociology," and more.
1988/752 pages/hardbound/illustrated/Instructor's Manual, Study Guide, Test Bank
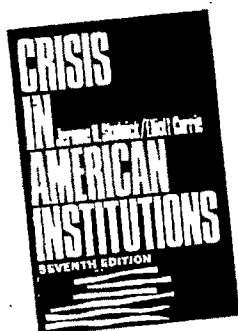
CRISIS IN AMERICAN INSTITUTIONS
Seventh Edition
Jerome H. Skolnick and Elliott Currie

These penetrating discussions of the most pressing social issues of our time provide a critical framework for understanding the nature of social problems. The Seventh Edition includes 13 new readings on such current issues as the tobacco industry lobby, hunger in America, teenage pregnancy, deregulation of the trucking industry, AIDS, and U.S. policy in Central America. The authors' incisive introductions have been updated throughout.
1988/608 pages/softbound/Instructor's Manual
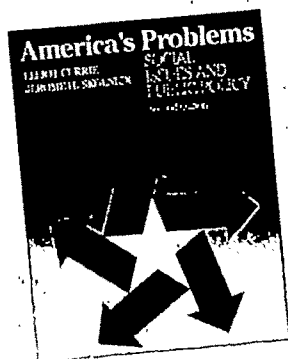
AMERICA'S PROBLEMS: Social Issues and Public Policy, *Second Edition*
Elliott Currie and Jerome H. Skolnick

This exceptional text helps students explore the broad social and economic structures and forces that shape current social problems, with special attention to social policy formulation and implementation. The new Second Edition has been thoroughly updated to reflect the most recent and sophisticated research and changes since 1984, including trends in inflation and oil prices, the new tax laws, and the trade deficit.
1988/608 pages/hardbound/illustrated/Instructor's Manual

VII

Social histories / Political scenes

Photo: Chicago Historical Society

CHICAGO '68
DAVID FARBER

An inside look at one of the epochal events of the 1960s—the 1968 Democratic Convention. Farber tells and retells the story of Chicago '68, in three different voices, from the perspectives of each of the major protagonists: the Yippies, the Mobilization to End the War in Vietnam, and Mayor Daley and the city's police.

"Sure to be a standard reference on the events of Chicago '68."—Assemblyman Tom Hayden, Santa Monica, California
Cloth $19.95   352 pages   24 halftones

WOMEN AND POLITICS
An International Perspective
·Second Edition
VICKY RANDALL

An excellent and concise introduction to, summary, and analysis of the enormous volume of writing on women and politics that has emerged since the 1960s. This second edition, greatly expanded and revised, provides the most comprehensive introduction available to the relationship between politics and women and is essential reading for students of politics, feminism, and the position of women in contemporary society.
Paper $12.95   368 pages
Library cloth edition $32.50

CONSTRUCTING THE POLITICAL SPECTACLE
MURRAY EDELMAN

Reinterpreting the contemporary political scene, Edelman shows how readily available political news bolsters existing inequalities and strengthens dominant ideologies.
Paper $7.95   144 pages
Library cloth edition $19.95

ECONOMIC BEHAVIOUR IN ADVERSITY
JACK HIRSHLEIFER

Economic Behaviour in Adversity brings together ten important essays, including several previously unpublished historical investigations and innovative theoretical analyses, on how individuals and societies function in times of disaster and conflict.
Cloth $35.00   320 pages   22 line drawings

IN PAPER

ORAL HISTORY AND DELINQUENCY
The Rhetoric of Criminology
JAMES BENNETT
$15.95   379 pages

MILITARY INSTITUTIONS AND COERCION IN DEVELOPING NATIONS
The Military in the
Political Development of New Nations, Expanded Edition
MORRIS JANOWITZ
$12.95   232 pages
Midway Reprint

XII

XIII

XIV

American Sociological Review

Volume 53 ● April 1988 ● Number 2

Official Journal of the American Sociological Association

SCOPE AND MISSION: The *American Sociological Review* publishes work of interest to the discipline in gneeral, new theoretical developments, results of research that advance our understanding of the most fundamental social process, and important methodological innovations. Like other publications of the Association, emphasis is given to exceptional quality. Unlike the more specialized journals of the Association, the *American Sociological Review's* primary focus is on whatever has the most general bearing on the knowledge of society.

Concerning manuscripts, address: Editor, *American Sociological Review,* Department of Sociology, Ohio State University, 300 Bricker Hall. 190 North Oval Mall, Columbus, Ohio 43210-1353. Phone (614) 292-6200.

(Articles in the *American Sociological Review* are indexed in the Social Sciences Index, Psychological Abstracts, Sociological Abstracts, Ayer's Guide, University Microfilms, Abstracts for Social Workers, International Political Science Abstracts, United States Political Science Documents, and Current Index to Journals in Education.)

MEMBERS OF THE COUNCIL, 1988

Officers of the Association

3

AMERICAN SOCIOLOGICAL REVIEW

Notice to Contributors

(Revised January 1987)

AUTHOR'S RESPONSIBILITY: *Manuscripts which do not adhere exactly to these suggestions will be immediately returned for revision. Only articles in the form specified below will be sent out for refereeing.*

The third edition of the *Publication Manual* of the American Psychological Association (available from A.P.A., P.O. Box 2710, Hyattsville, MD 20784) is very helpful in clarifying headings, abbreviations, table preparation, and the like.

Since manuscripts are evaluated as anonymous submissions, please keep identifying material out of the manuscript. Attach a cover page giving authorship, institutional affiliation, acknowledgments, and the date the article is submitted. Provide only the title as identification on the manuscript and abstract. Every effort should be made by the author to see that the manuscript itself contains no clues as to the author's identity.

Submission and Preparation of Manuscripts

1. Submit five (4) copies and retain the original for your files. Copies may be Xerox, mimeograph, or multilith, but not carbons. Manuscripts must be accompanied by a $10.00 manuscript submission fee, payable to the American Sociological Association.

2. Enclose a stamped self-addressed postcard for acknowledgment of manuscript receipt. Since manuscripts will not be returned by reviewers to the editorial office, the editorial office will not return them to authors.

3. All copy must be *typed, doublespaced* (including indented material, footnotes, and references) on 8½ by 11 inch white opaque paper. Line must not exceed six (6) inches. Margins must be a minimum of 1¼ inch on left and foot.

4. Include four (4) copies of an abstract of *no more* than 150 words.

5. Tables should be numbered consecutively throughout the article and typed on separate sheets at the end of the manuscript. Insert a location note at the appropriate place in the text, e.g., "Table 2 about here." Each table must include a descriptive title and headings to columns. Gather general footnotes to tables as *"Note:"* or *"Notes:,"* and use *a, b, c,* etc., for specific footnotes. Asterisks * and/or ** indicate significance at the 5 percent and 1 percent levels, respectively.

6. Illustrations submitted with the final draft must be of professional quality, ready for reproduction, executed on white paper or vellum, in black ink, with clear, medium weight, black lines and figures. All lettering on figures should be executed by an artist in pen and ink, by means of dry transfer letters, or by applying typeset material to the prepared artwork. Typewritten lettering should not appear in illustrations. Figures should be capable of legible reduction to a width size no larger than 5³⁄₁₆ inch (full page) or 2⁹⁄₁₆ inch (one column); they should be numbered consecutively, and the number and author's name should be penciled lightly on the back of each. All illustrations must have captions, which should not appear on the artwork but should be typed, double-spaced, on a sheet at the end of the manuscript.

7. Equations must be typed and important displayed equations identified by consecutive arabic numbers in parentheses on the right. Expressions should be aligned and compound subscripts and superscripts clearly marked if there is any potential for confusion. Indicate bold face characters by drawing a wavy line (⌄) under them; a single underline (⌄) means italic to a printer. Clarify all symbols with notes in the margins of the manuscript. Circle these and all other explanatory notes not intended for printing.

8. Three kinds of footnotes are possible each serving a different purpose:

    a. *Content footnotes.* Content footnotes are explanations or amplifications of the text. Because they are distracting to readers and *expensive* to include in printed material, an author should include important information in the text and omit irrelevant information. *Content footnotes will not be allowed generally.*

    Rather than footnoting long or complicated material, such as proofs or derivations unnecessary to the text, consider (i) indicating in a *short* footnote that the material is available from the author, (ii) depositing the material in a national retrieval center and including an appropriate footnote, or (iii) adding an appendix. If an appendix is used, the reference in text should read, for example: "(see Appendix A for complete derivation)."

    Text footnotes should be numbered consecutively throughout the article with superscript arabic

numerals. If after a footnote occurs it is later mentioned, use a parenthetical note "(see note 3)," rather than the superscript number.

b. *Reference footnotes*. Footnotes are used for reference purposes only to cite material of limited availability. Acceptable reference footnotes include (i) legal citations, which should follow the footnote style of *A Uniform System of Citation* (1967) published by the Harvard Law Review Association, (ii) copyright permission footnotes, (iii) unpublished works and works in progress.

c. *Table footnotes*. Table footnotes are appended *only* to a specific table. Footnotes to a table should be *lettered* consecutively *within* each table with superscript lowercase letters. (See 5.)

9. Acknowledgments, credits, and grant numbers are placed on the title page with an asterisk.

Reference Format

A. *In the text:* All source references are to be identified at the appropriate point in the text by the last name of the author, year of publication, and pagination where needed. Identify subsequent citations of the same source in the same way as the first. Examples follow:

1. If author's name is in the text, follow it with year in parentheses ["... Duncan (1959) ..."].

2. If author's name is not in the text, insert in parentheses the last name and year ["... (Gouldner 1963) ..."].

3. Pagination follows year of publication after a comma ["... Kuhn (1970, p. 71)."].

4. Give both last names for dual authors. Give all last names on first citation in text for more than two authors; thereafter use "et al." in the text. When two authors have the same last names, include initials in the text. For institutional authorship, supply minimum identification from the beginning of the complete citation ["... U.S. Bureau of the Census 1963, p. 117) ..."].

5. Separate a series of references with semicolons and enclose them within a single pair of parentheses ["... (Burgess 1968; Marwell et al. 1971, pp. 386–87; Cohen 1962) ..."].

B. *In the appendix:* List all items alphabetically by author and, within author, by year of publication in an appendix titled "REFERENCES." The reference appendix must be complete and include all references in the text. The use of "et al." is not acceptable in the appendix; list the names of all authors using full first names. (See A.4. for text format.)

If there is more than one reference to the same author and year, distinguish them by the letters a, b, etc. added to the year ["... Levy 1965a, p. 331) ..."].

The first letter of each word in an article title should be capitalized. Titles of books and journals are printed in italics, so each word of the title should be underlined.

Give the publisher's name in as brief a form as is fully intelligible. For example, John A. Wiley and Sons should be "Wiley."

If the cited material is unpublished, use "forthcoming" with name of journal or publisher; otherwise use "unpublished."

Examples follow:

1. *Books:*

Mason, Karen O. 1974. *Women's Labor Force Participation and Fertility.* Research Triangle Park, NC: National Institutes of Health.

U.S. Bureau of the Census. 1960. *Characteristics of Population.* Vol. 1. Washington, D.C.: U.S. Government Printing Office.

2. *Periodicals:*

Conger, Rand D. Forthcoming. "The Effects of Positive Feedback on Direction and Amount of Verbalization in a Social Setting." *Pacific Sociological Review.*

Goodman, Leo A. 1974a. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.

———. 1974b. "The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I—A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–1259.

3. *Collections:*

Clausen, John A. 1972. "The Life Course of Individuals." Pp. 457–514 in *Aging and Society,* vol. 3, *A Sociology of Age Stratification,* edited by M.W. Riley, M. Johnson, and A. Foner. New York: Russell Sage.

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165–90 in *Annual Review of Sociology,* vol. 1, edited by A. Inkeles, J. Coleman, and N. Smelser. Palo Alto, CA: Annual Reviews.

See 1986 and later issues for further examples.

ANNOUNCEMENT

Beginning July 1, 1988, manuscript submission fees for all ASA journals will be increased to $15.

CONTRIBUTORS TO THIS ISSUE

■ GERHARD LENSKI (Rethinking Macrosociological Theory) is Alumni Distinguished Professor of Sociology at the University of North Carolina at Chapel Hill. He recently completed an article for the (Italian) *Encyclopaedia of the Social Sciences* on the development of societal taxonomies from ancient times to the present. He is currently working on a monograph tentatively titled *Ecological-Evolutionary Theory: Principles and Applications*, which he will summarize at a conference of the Hungarian Sociological Association this month.

■ ALAN NEUSTADTL (Corporate Political Groupings: Does Ideology Unify Business Political Behavior?) is Assistant Professor of Sociology, University of Maryland, College Park. He studies corporate political behavior, interest group politics, and networks of political action. DAN CLAWSON is Associate Professor of Sociology, University of Massachusetts at Amherst. He is critically analyzing Quality of Work Life programs. Together they are studying the role of corporate groupings on the political realignment of the 1970s and 1980s and have published on related topics.

■ DAVID JAMES (The Transformation of the Southern Racial State: Class and Race Determinants of Local-State Structures) is Assistant Professor of Sociology, Indiana University, Bloomington. His current research extends his examination of the causes and effects of local-state structures. He is investigating the extent to which central city and suburban school systems interact to increase public school segregation in metropolitan areas. He is also collaborating with Pamela Barnhouse Walters on a study of enrollments and racial inequalities in public schooling in rural areas of the South between 1890 and 1980.

■ LAUREL RICHARDSON (Secrecy and Status: The Social Construction of Forbidden Relationships) is Professor of Sociology, Department of Sociology, The Ohio State University. She is the author of *The Dynamics of Sex and Gender*, 1988, and of *The New Other Woman*, published in 1985 with the English language paperback and German and Japanese translations released in 1987. In addition to her continuing interest in gender, relationships, and theory, she is currently analyzing sociological narrative strategies.

■ BARRY MARKOVSKY (Power Relations in Exchange Networks) is Assistant Professor of Sociology at the University of Iowa. In addition to his research in the areas of social networks and exchange, recent papers in the *American Sociological Review, Social Forces, Sociological Theory,* and elsewhere address status processes, distributive justice, integrating micro and macro theories, and the interplay of social psychological and physiological processes. He is co-editor (with Edward J. Lawler) of the annual volume *Advances in Group Processes.* DAVID WILLER, Professor of Sociology, University of Kansas, is currently using elementary theory to develop micro-micro linkages. This research seeks to bring experimental research to bear on historical explanations. He is also studying the influence of types of network connection on the distribution of power in social structures. His most recent book, *Theory and the Experimental Investigation of Social Structures,* was published in 1987 by Gordon and Breach. TRAVIS PATTON is a Ph.D. candidate in the Department of Sociology, University of Kansas. In his dissertation, he is constructing a formalized concept of power and applying it to some classical issues in sociological theory.

■ WAYNE J. VILLEMEZ (When Bigger is Better: Differences in Individual-Level Effect of Firm Size and Establishment Size), Professor and Chair, Department of Sociology, Louisiana State University, is working on a monograph with William Bridges on the individual-level effects of bureaucratic organizational forms. He continues to work with colleagues at LSU on the process of wage determination in new industries and on household network structures. WILLIAM P. BRIDGES is Associate Professor of Sociology at the University of Illinois at Chicago. He is continuing his research and writing on the interaction of competitive labor markets and organizational personnel practices and their joint effects on wages and labor mobility.

■ MOSHE SEMYONOV (Bi-Ethnic Labor Markets, Mono-Ethnic Labor Markets, and

Socioeconomic Inequality) is Associate Professor and Director of the Institute for Social Research, Department of Sociology, Tel Aviv University. His research interests lie in the area of comparative stratification systems and in structural sources of socioeconomic inequality. Current research focuses on change and stability in the social organization of occupational and local labor markets. Recent research on the Israeli labor market was published (ILR Press, 1987) in a book entitled *Hewers of Wood and Drawers of Water: Noncitizen Arabs in the Israeli Labor Market* (co-authored with Noah Lewin-Epstein).

■ CLIFFORD C. CLOGG (A Flexible Procedure for Adjusting Rates and Proportions Including Statistical Methods for Group Comparisons) is Professor of Sociology and Statistics and Senior Research Associate of the Population Issues Research Center at the Pennsylvania State University. He is currently the editor of *Sociological Methodology*. His research interests include the social demography of the labor force, unemployment and underemployment, and statistical models for categorical data. SCOTT R. ELIASON is a graduate student in the Department of Sociology at the Pennsylvania State University. His research interests include the sociology of labor markets, careers and the life cycle, and quantitative methodology.

■ JOHN M. HEINEKE (Crime Deterrence and Choice) is Professor of Economics in the Department of Economics, Santa Clara University. His long-term research interests in the area of rational decision making in uncertain environments underpins his interest in modeling criminal behavior. His book, *Economic Model of Criminal Behaviour,* is a series of

papers devoted both to theoretical models of criminal behavior and to econometric testing of such models. Currently his research efforts are directed to the problem of specifying conditions under which it is possible to aggregate decisions taken by heterogeneous individuals into a consistent aggregate decision function.

■ ROSS L. MATSUEDA (Economic Assumptions vs. Empirical Choice) is Assistant Professor of Sociology at the University of Wisconsin-Madison. In addition to his research in criminal decision-making (with Piliavin and Gartner), he is developing and testing a structural symbolic interactionist theory of delinquency (with Karen Heimer), estimating sibling models of delinquent behavior, and continuing his research on statistical power in covariance structure models (with William T. Bielby). IRVING PILIAVIN is Professor of Social Work and Sociology at the University of Wisconsin. He is currently involved in two lines of inquiry in addition to his work with Matsueda and Gartner. The first involves the study of phenomena associated with entry into delinquent activity among adolescents and pre-adolescents; the second concerns the factors leading to long-term homelessness and exits from homelessness. ROSEMARY GARTNER is Assistant Professor of Sociology at the University of Iowa. She is continuing her work with Piliavin and Matsueda, as well as examining cross-national and temporal variation in the sex differential in homicide victimization. She co-authored, with Dane Archer, *Violence and Crime in Cross-National Perspective* (Yale University Press, 1984), which received the ASA Criminology Section's 1986 Award for Distinguished Scholarship.

ERRATA

The following is a correction of a printer's error in "Gender and Promotion in Segmented Job Ladder Systems," by Thomas A. DiPrete and Whitman T. Soule (*ASR* 53:35).

In Table 4, under column 2, for the variable grade 12, the number 0.32 (under −0.53) should be deleted.

ANNUAL EDITOR'S REPORT, 1987

In 1987, 425 manuscripts were submitted to *ASR*, about the same number as in 1984, but down from 448 in 1986. About 17 percent of the submitted manuscripts were eventually published; 65 percent were rejected outright; and revise and resubmitted manuscripts currently in process total 18 percent. Editorial lag, or date from receipt of manuscript to data of decision, was 53 days, a record low. Publication lag, or date from manuscript acceptance to manuscript publication, was 4 months. *ASR* published 53 articles, 12 research notes, and 7 comments and replies. In addition, *ASR*, as the ASA journal of record, added three new features: table of contents of ASA journals and publications, list of ASA members whose deaths were reported the previous year, and guest editorials. The increase in published items within the same page allocation was achieved by slightly reducing the length of manuscripts.

Manuscripts were classified according to the subject matter of their main dependent variable. Areas that had more than 10 manuscripts included political, 40; criminology, 35; stratification-mobility, 34; sex and gender, 29; social psychology, 23; sociological thought and theory, 21; methodology and work, each 20; race-ethnicity, 18; development-world systems, 17; demography, 16; community, 15; social organization and marriage and family, each 11. A sample of one-half of the manuscripts submitted revealed that only 20 percent of the "contact authors" were women. Of the 505 reviewers used during the year, 27 percent were women.

In collaboration with other journals, *ASR* encouraged authors of accepted papers to document the sources of their data more fully in the Reference section of their papers. In addition, authors were asked whether their data sets were publicly available or deposited at the Institute of Political and Social Research at the University of Michigan. The cover stock of *ASR* was improved as was its formatting. Beginning with the April issue, as part of the long term ASA policy to restore the physical quality of its journals, the type size of *ASR* will be larger.

ASR is deeply indebted to the Associate Editors for assuming the heavy burden of evaluating manuscripts and giving their colleagues detailed suggestions on how to improve their manuscripts. Those who have served their three-year terms include Sarah Fenstermaker Berk, Diane H. Felmlee, Mary R. Jackman, Gary F. Jensen, William R. Kelly, John R. Logan, Francois Nielsen, Barry Schwartz, and Judith Seltzer. *ASR* welcomes its newly appointed Associate Editors to three years of hard work: William C. Bailey, Charles Camic, Arlene K. Daniels, Guillermina Jasso, Barrett A. Lee, Toby Lee Parcel, Richard B. Rubinson, and Gaye Tuchman. I am especially grateful to the Deputy Editors, Craig J. Jenkins and Robert R. Kaufman, for their general assistance in running the journal. Kummi Ranjit, our former Managing Editor, left *ASR* to become a reporter of Indian contemporary affairs. The editors are grateful for her excellent service to *ASR* and welcome Mary Lee Raines, the new Managing Editor, who has skillfully organized us to work even harder to produce a high quality journal.

Readers have responded to my editorial statement that *ASR* is open to a wide range of manuscripts. I hope that the published articles will increasingly reflect this openness. Continue to send in your best contributions. We will do the work.

William Form

RETHINKING MACROSOCIOLOGICAL THEORY*

GERHARD LENSKI

University of North Carolina at Chapel Hill

Macrosociologists have too long been complacent about the state and quality of their theory. If macrosociological theory is to progress, and not merely proliferate, several major changes are needed in the way theory is usually formulated. Above all, theories must become falsifiable. This will require that key concepts be stated in an operationally unambiguous manner and that relationships among variables also be specified unambiguously. In addition, theories ought explicitly to incorporate biological and environmental constants and not leave them implicit or ignore them, as is too often the case. Finally, theories should be multilayered, with a general theory employing a covering principle to identify a key variable or variables on which a series of derivative special theories can be based. Above all, theory construction requires greater rigor and discipline than is currently the norm.

Sociologists and other social scientists are usually irritated by the refusal of many in the natural sciences to acknowledge the scientific nature of their disciplines. Sometimes this is thought to be a not too subtle form of intellectual snobbery, or a technique for reducing competition for scarce resources. It is also viewed as the product of an inadequate conception of the nature of science itself. Science, it is argued, is basically a mode of inquiry and social scientists employing the scientific method deserve the same respect and support as those who work in the natural sciences.

This view of science and the social sciences is not entirely without merit, but it neglects one very important consideration. Science is more than method: *its ultimate aim is the development of a body of "verified" general theory.*[1] Thus, for many who are trained in

the natural sciences, the acid test of the scientific status of a discipline is the quality of its theory, and, regrettably, most social science theories in general, and macrosociological theories in particular, do not fare well by this standard—especially those that the majority of social scientists themselves acclaim most highly.

Macrosociological theories fall short when judged by at least two important criteria. First, most of these theories are not falsifiable in the same unambiguous manner as theories in the natural sciences. Second, unlike theories in the natural sciences, macrosociological theories lack substantive conceptual links to established theories in other scientific disciplines.[2] These are critical flaws from the

[1] I have put the word "verified" in quotation marks, since nomothetic theory can never be verified in the sense that it is proven beyond any possibility of subsequent refutation. Scientific theories are considered tentatively "verified," however, when repeated empirical tests yield results consistent with them and when competing theories

have been proven to be in error. Furthermore, "verified" theory is a goal that the sciences strive to attain, and it is progress toward this goal that is essential, not its actual attainment.

I have described the aim of the sciences as the development of a body of *general* theory to indicate that the aim goes beyond the formulation of scattered and disparate partial or limited theories. *Webster's Third New International Dictionary* defines theory in this sense as "the coherent set of hypothetical, conceptual, and pragmatic principles forming the general frame of reference for a field of inquiry." For the discipline of sociology, I would define general theory as "a coherent set of principles capable of providing a general frame of reference for the discipline as a whole or some major segment (e.g., macrosociology, social psychology) of it."

[2] Within the natural sciences, there are important substantive links between theories in the various disciplines. One can move from subatomic physics to evolutionary ecology and back again without a major break in the web of theories (e.g., evolutionary theory in biology incorporates certain fundamental principles from chemistry, just as evolutionary theory in chemistry incorporates certain fundamental principles from phys-

standpoint of those trained in the natural sciences and they necessarily engender skepticism about the entire enterprise in which social scientists are engaged.

For social scientists, the question that must be asked is whether these are necessary and unavoidable features of theory in our disciplines or whether they are flaws that we have negligently allowed to survive and flourish at great cost to the development and intellectual health of our disciplines. I am not so presumptuous to claim that I have the definitive answer to this question, or even that I have a clear understanding of all the issues involved. I am convinced, however, that the problem needs to be discussed more openly than is currently the case and that such a discussion could yield substantial benefits for all the social sciences and, perhaps ultimately, for our understanding of human societies and the problems that confront them.

I realize, of course, that not all sociologists accept the idea that sociology is (or should be or can become) a science or that it should be guided by the standards of science. Those who reject this view of the discipline will find little of interest in this paper, since it is premised on the belief that without such a commitment it is difficult to justify public support of sociology.

SOME OBJECTIONS CONSIDERED

The problem I have raised is not new. Many of the issues have been discussed before, and seemingly cogent objections raised to proposals for employing the natural sciences as models for the social sciences. These objections need to be considered at the outset, since many insist that is it folly for social scientists to look to the natural sciences for guidance.

Probably the most persuasive objection to proposals for incorporating standards developed in the natural sciences in the social sciences is based on the obvious fact that "people are different" from the objects of study in the natural sciences. People possess consciousness and wills. Subatomic particles, chemical elements, stellar systems, cells, and even many of the organisms studied by biologists do not. Furthermore, human behav-

ior is profoundly influenced by extraordinarily complex cultural traditions, a fact that sets us apart from even that part of the biotic world that shares with us the attributes of consciousness and will. As a result, the determinants of human behavior are extraordinarily complex—far more complex apparently than the determinants of any of the objects studied by natural scientists. Add to all of this the fact that the study of many of the most important forms of human behavior—including, especially, the behavior of large scale social systems—are not amenable to controlled experimentation under laboratory conditions, and it is clear that the social sciences cannot operate in the same way, or be governed by the same rules, as the natural sciences.

There is, of course, much truth to this assertion. The social sciences must, of necessity, adapt their mode of inquiry to these realities. But this does not justify the abandonment or neglect of all the standards of modern science. This is readily recognized with respect to certain standards governing research and to certain principles of methodology. Its relevance for theory, however, seems not to be so evident.

Another frequent objection to proposals that the social sciences might benefit from closer adherence to certain principles and practices developed in the natural sciences is that it is demeaning for the social sciences to ape the natural sciences. This argument makes an obvious appeal to tribal or sectarian pride: we are just as clever and creative as they, and we should create our own customs and standards.

Such an appeal has considerable force for reasons that C. P. Snow (1959) made clear in his insightful "two cultures" lectures a generation ago. Those who have been raised in the humanist cultural tradition—as have the vast majority of social scientists—not only have little understanding or appreciation of the techno-scientific tradition, but tend to be fearful of it and hostile toward it.[3]

---

ics—though not in a reductionist manner). In contrast, most macrosociological theories ignore, deny, or minimize the relevance of natural science theory.

[3] It is also true, as Snow pointed out, that those raised in the techno-scientific cultural tradition have little understanding or appreciation of the humanist tradition and tend to be disdainful of it, and this greatly exacerbates the problem. In Snow's opinion, we need a new generation of scholars and decision-makers who have sufficient training in both traditions to appreciate and benefit from both. In a later edition of the volume containing his "two cultures" lecture, Snow added an Epilogue in which he expressed the hope that the social

Objections such as those noted above are not sufficiently persuasive, however, given the seriousness of the issues involved. For even if we ignore the criticisms of colleagues in the natural sciences, we need only reflect on the muddled state of theory in our field to recognize the problem confronting us. Above all, we are faced today with a vast and constantly growing array of theories. While it is true that many older theories are now largely ignored, this is not because they have been tested rigorously and found wanting. Usually it is only because newer theories have come along and crowded them out of our limited span of attention—either because of vigorous promotion by energetic proponents or simply because of their novelty appeal.

One might even say that in the social sciences, theory has not advanced nearly so much as it has proliferated. In contrast, one of the most basic attributes of theory in the natural sciences is that it is continuously being refined and improved. Theories stimulate research, and research leads to theoretical advances. While it is true that there are some instances of this in macrosociology, this kind of fruitful interaction has not been common. Much of the blame for this lies with our theories and the theorists responsible for them.

One measure of the present state of theory in sociology is the courses on theory that we offer graduate students. The content of such courses is usually a sample of the writings of the Holy Trinity—Marx, Weber, and Durkheim—plus more limited attention to the work of others, such as Simmel, Mead, Parsons, Giddens, Habermas, and Goffman. Emphasis in theory courses usually seems to be on exegesis and/or a critique of texts rather than on the relevance of theory for research.[4] In this respect, the teaching of sociological theory has more in common with seminary instruction in theology and biblical studies

than with graduate instruction in the natural sciences. By teaching theory for its own sake, rather than in conjunction with a distinctive research tradition, we tend to raise up new generations of theorists who too often emulate the old masters in ways that hinder, rather than advance, the cause of macrosociology.

What is the solution, or is there a solution?

I believe that there is a solution, but it will require substantial changes in the ways in which macrolevel theory is usually formulated.

GUIDANCE FROM THE PAST

Thirty years ago, a faculty group at Cornell University, who described themselves as having been "both puzzled and stimulated" by the writings of Talcott Parsons, met regularly during the academic year to discuss his work (Black 1961, p. v). This led to the publication several years later of a collection of essays titled *The Social Theories of Talcott Parsons: A Critical Examination.*

It may not be coincidental that the most devastating critique in this collection was by a nonsociologist, Max Black. In an essay titled, "Some Questions About Parsons' Theory," the most troubling issues he raised concerned Parsons' concepts.

While absolving Parsons of the charge that his concepts were too abstract or remote from experience, Black (1961, p. 281) was highly critical of what he regarded as "an endemic ambiguity of scope in Parsons' use of key terms," an ambiguity that manifested itself in "a constant vacillation between narrow and broad" usages of key concepts. Black also criticized the frequent redefinition of concepts in ways that he found to be inconsistent. In the case of the crucial pattern variables, the definitions changed so often that Black (p. 284) wrote, "I think it would not be unfair to call them 'chameleon concepts.' "

Black (p. 283) attributed this flaw in Parsons' work to the fact that the basic conceptual scheme was not adequately linked to empirical generalizations and theories, creating what he called "disturbing consequences . . . for Parsons' claim to have provided a *scientific* framework for the social sciences" (emphasis in the original).

As Black explained, "The supreme virtue of a *scientific* classification, whether in physics, chemistry, or biology, is that it arises from, and is in some sense demanded by, a system

---

sciences might serve as a bridge between the humanist and techno-scientific cultural traditions. The fulfillment of that hope, however, has yet to be achieved.

[4] I am not suggesting that sociology abandon the study of prototheorists whose works fall short of the standards of modern scientific theory. Their works often contain invaluable insights that must not be lost. In directing graduate students to read such works we should encourage them to search out these insights and think about how they can be incorporated into more rigorous theoretical formulations. We should not, however, leave the impression that these older works conform to the standards of good scientific theory.

of well established empirical generalizations and theories." As Black observed, this was true of Mendeleev's classification of the elements and of Linnaeus' classification of species, and, he might have added, as a notable example from the social sciences, of Thomsen's classification of prehistoric societies and their cultures (Harris 1968, p. 146; Woodbury 1968, p. 379), which became the foundation for all subsequent theory and theory-directed research in archaeology. But this is not true of Parsons' pattern variables, or of comparable concepts of many other currently more fashionable macrosociological theorists.

Classificatory schemes and the theories built on them that do not grow out of, build on, and interact with an established body of empirical research, tend to be sterile, in the sense that they seldom generate empirical research. As a consequence, they fail to develop the intellectual resources required for cumulative growth and continuing refinement of their formulations.

Unfortunately, much of what passes for theory in macrosociology today has these defects. This is why the field is plagued with a multiplicity of competing theories: too few of them can be eliminated because too few are truly *falsifiable* (Popper 1959). And this is why theory is so often taught as the "history of theories" (and/or theorists), rather than as the theory of human societies, or the theory of social systems. Thus, sociological theory—especially macrolevel theory—is more an art form than a science.

MINIMAL STANDARDS: UNAMBIGUOUS CONCEPTS, SPECIFIED RELATIONSHIPS, AND FALSIFIABLE THEORY

For those who believe that the ultimate aim of every science is the development of a body of "verified" general theory, an essential means to that end is the creation of *falsifiable* theory—theory that can, either in whole or in part, be tested empirically and demonstrated to be false, if, indeed, it is false.[5]

While most sociologists would not challenge this assertion, and even regard it as platitudinous, its implications are commonly ignored. For example, there are at least two requirements of falsifiable theory that are still widely ignored by many leading macrosociological theorists. First, key variables must be defined in an operationally unambiguous manner. Second, *relationships* among and between variables must be specified unambiguously. Unless these conditions are met, there is no way for the community of scholars to decide and agree if tests of a theory support or contradict it, and thus no way to rid the discipline of unsatisfactory theories or to refine those that are promising.

When the implications of falsifiable theory are spelled out in this way, it can be shown that much that passes for theory in sociology today fails to meet even minimum standards. As Black's critique of Parsons' work made clear, Parsons' concepts were both ambiguous and internally inconsistent. Unfortunately, Parsons has not been alone in his neglect or lack of awareness of the requirements of falsifiable theory. Many contemporary macro-level theorists share these traits, as do also the old masters—Marx, Weber, and Durkheim—whom we so often hold up before students as exemplars to be emulated. It is no accident, for example, that Marx's writings have generated almost as many divergent interpretations as the Bible. While this is a source of great strength from an ideological standpoint, since the fuzziness of concepts and internal contradictions in texts make falsification all but impossible and provide almost limitless possibilities for reinterpretation as this becomes necessary, good ideology is not the same as good science.

At a minimum, it should be possible for theorists to diagram the causal relationships among concepts hypothesized by their theories. Whatever its limitations and flaws, Blau and Duncan's *American Occupational Structure* (1967) is to be commended for the precision and clarity with which its basic theoretical argument was set forth. The five variables in the basic model (Father's education, father's occupation, respondent's educations, respondent's first job, and respondent's occupation in 1962), unlike Parsons' variables, all were grounded in an established

---

[5] Most macrolevel theories in the sciences do not lend themselves to single definitive tests of the entire theory. Most tests pertain only to selected aspects. When the results of such tests contradict what otherwise appears to be a sound theory, this leads to modifications of, or adjustments in, the theory to accommodate the new information. This is, of course, the basis for the process of theoretical refinement and advance that has been such

an important and distinctive feature of the natural sciences.

research tradition and therefore were reasonably unambiguous (father's occupation and respondent's first job, however, left something to be desired as the authors acknowledged). Furthermore, the hypothesized relationships among the variables were all clearly specified so that falsifying the theory would have been a simple matter if the theory were false. Finally, the clarity of the model has enabled critics to identify its limitations (i.e., discover what the model is not able to explain).

What has emerged from tests of the theory by Blau and Duncan themselves and by others is that (a) greater specification of some of the variables is needed, and (b) the model is incomplete, and other variables must be taken into account.[6] Thus, while the original statement of the theory was flawed or imperfect, it was readily amenable to correction and refinement—something that has been much more difficult, or impossible, to achieve in most macrosociological theories.

To cite the Blau and Duncan theory of status attainment or other comparably specified theories is to invite the rejoinder that such specification is only possible when large amounts of quantitative data are available. As a corollary, it is often asserted that this kind of specification is impossible in many, or even most, of the important kinds of macrolevel theories in which cross-national and historical materials provide the data base. There are too few cases for rigorous statistical analysis and too much missing data. Therefore, presumably, we need not bother to construct the same kinds of precisely specified theoretical models when working with cross-national and historical materials as when working with sample survey or census data.

Once again, there is an element of truth to the argument. The kinds of statistical analysis that are possible in the study of a subject such as status attainment are not possible in the

study of patterns of inequality in human societies from prehistoric times to the present. This does not mean, however, that theories of the latter need not be falsifiable or that concepts and their relationships need not be stated clearly and unambiguously. It only means that the tasks of theorists and researchers are more difficult and the pitfalls more numerous.

At a minimum, all theories should be presented diagrammatically, as in path analyses.[7] This would compel theorists to think out the nature of relationships among their variables in a way that is not necessary when one presents a "theory" in the traditional manner. As Parsons' unhappy experience demonstrates, purely verbal formulations of theory too easily allow a theorist to waffle and shift ground subtly in ways that even the theorist may not recognize and readers have great difficulty in detecting.

There are within the field of stratification a number of instances of the kind of presentation of theory I have in mind. *Power and Privilege: A Theory of Social Stratification* (Lenski 1966, pp. 89, 439) is an early approximation. Janet Saltzman Chafetz's *Sex and Advantage: A Comparative Macro-Structural Theory of Sex Stratification* (1984, p. 11) and Jonathan H. Turner's *Societal Stratification: A Theoretical Analysis* (1984) are two more recent examples. In the first two instances, the variables were consciously drawn from established traditions of empirical research. Turner's work, by comparison, is notable for the special effort made to specify the form of relationships among variables.

Although there are errors and deficiencies in all three of these theories, the way they are stated facilitates efforts to identify and correct them. The authors cannot hide confused or mistaken views behind verbal smoke screens. Thus, opportunities for testing and then refining and improving these theories (assuming they have some merit) are substantially greater than for stratification theories that have not been formulated in the same

---

[6] Critics have also shown that the theory is far from being a general theory of social stratification. For example, while it explains a considerable part of the process by which individuals are distributed among various occupations in modern industrial societies, it does not explain how different occupations have come to be associated with different levels of rewards in various societies, or in a single society at various times in its history. Also, it provides few clues to the causes of changes in the status attainment process itself. But once again, the clarity of the model makes it easy to recognize these limitations.

---

[7] Alternatively, theories may be expressed even more precisely in the form of algebraic equations, since they allow for more precise specification of relationships among variables. This level of precision is not imperative, however, in much of today's macrosociology, although it may be that if theory construction can be improved sufficiently in the years ahead, we will arrive at the point where algebraic formulations of most theories become routine.

way. One might contrast them in this regard with Talcott Parsons' (1953) "Revised Analytical Approach to the Theory of Social Stratification," or Anthony Giddens' (1973) more recent *The Class Structure of Advanced Societies*, neither of which readily lends itself to unambiguous falsification.[8]

In summary, scientific standards concerning theory formulation must not be restricted merely to theories that lend themselves to rigorous quantitative tests. They can *and should* be applied to theories for which the quality of the data are much less satisfactory. In such instances, the standards governing theory formulation must not be compromised merely because the data are flawed.[9] Variables, and relationships among them, still must be specified clearly and unambiguously so that tests of the theory may be as rigorous as the data allow.

ADDITIONAL DESIDERATA: MULTILAYERED THEORY AND THE INCLUSION OF CONSTANTS

Beyond the adoption of these minimal standards, there are other changes that would improve the quality of macrosociological theory. Two that seem especially important are the development of multilayered theories and the inclusion of constants.

A multilayered theory is one in which a broadly inclusive general theory established a covering principle or covering law (Weingartner 1967) from which a series of more limited special theories can be derived. For example, if a general theory hypothesizes that categoric differences in X are the primary causes of variations in a series of other societal variables, Y_1, Y_2, Y_3 . . . Y_n, then this principle can be used to generate a series of special theories concerning various subsets of societies, the members of which are similar

with respect to X. To illustrate, if the theory is concerned with all human societies, past and present, and if X is the mode of production in societies, then one can derive from the general theory a series of testable special theories for each of the sets of societies defined by the various modes of production.

The field of economics provides a good example of the use of multilayered theory. It has a general theory of markets that asserts that markets vary primarily in response to differences in the numbers and relative market strength of buyers and sellers. This general proposition establishes a foundation for a series of special theories dealing with monopolistic markets, monopsonistic markets, oligopolistic markets, oligopsonistic markets, and so forth.

Such a theory, or family of theories, is not likely to be deterministic, since it normally involves populations that are subject to many more influences than those dealt with in the theory and in available data. Therefore, predictions concerning the values of the dependent variables must necessarily be cast in probabilistic terms.[10]

Multilayered theory can provide a way of overcoming an otherwise legitimate objection to proposals that theoretical arguments should be presented diagrammatically as well as verbally. If a theory is complex, as most macrosociological theories are, efforts to diagram it in its entirety within a single causal model may become visually confusing and, thus, potentially counterproductive. If, however, a theorist is willing to assign priorities to independent variables, identifying those that are most powerful, it then becomes possible to limit the general theory to a statement of the covering principle governing these priorities. More detailed statements of causal relations can then be divided among the special theories derived from that general theory.

Evolutionary ecology (Lenski and Lenski 1987; Lenski, unpublished) provides an

---

[8] Judging from a statement in the preface to his more recent volume, *The Constitution of Society: Outline of the Theory of Structuration*, Giddens (1984) has no desire to conform to scientific standards of theory construction. In his words, "The task of constructing sets of stably established generalizations, which is (perhaps) the lynchpin of the endeavors of the natural sciences, is not an ambition of much relevance to social science. Or so I propose." (p. ix)

[9] If anything, the need to maintain standards is even more important when data are flawed. When rigorous tests of theories are impossible, ambiguity in concepts and specifications of their relationships is harder to detect.

[10] Logically, it should be possible to extend this process of employing principles from more general statements of theory as a basis for constructing special theories all the way to the point where, by a process of adding more and more constraining or limiting principles, one arrives at deterministic predictions for individual societies. This would be possible, however, only if one were able to incorporate into the theory, or family of theories, all of the relevant variables. This is clearly impossible in the study of human societies and other macrosociological phenomena.

example of the kind of multilayered theory I am suggesting and also illustrates the potential for refinement this kind of theory contains. In the earliest version of this theory, presented in *Power and Privilege* (1966), I hypothesized that variations in subsistence technology have been the most powerful single cause of variations in societal systems of stratification in the total universe of human societies, past and present. This proposition became the basis for a series of special theories dealing with various subsets of societies classified on the basis of their dominant form of subsistence technology (e.g., hunting and gathering societies, simple horticultural societies, advanced horticultural societies, agrarian societies, industrial societies).

Within that framework, I was sometimes able to develop a third tier of theories by adding other independent variables (as indicated by the general theory) to certain of the special theories. For example, in the discussion of advanced horticultural societies, I noted (Lenski 1966, pp. 160ff.) that the general theory hypothesized that the biophysical environment can be a cause of variations in political development and inequality, and then went on to show that advanced horticultural societies in tropical rain forest regions have been politically less developed and more egalitarian than advanced horticultural societies in other kinds of environments.

For present purposes, however, it is the structure of theories that is important, not their content. The content may be flawed in a number of respects, but if the structure of the theory and its basic ideas are sound (i.e., if the minimum standards for theory construction cited previously have been met and the essential ideas have merit), then it should be possible, over time, through the interplay of theory and research, gradually to reduce the number of substantive errors in the theory and move a few steps closer to "verified" theory.

With a well-designed, multilayered theory, or family of theories, it should also be possible over time to develop a coherent set of principles capable of providing a meaningful framework for macrosociology both for teaching and research (as evolutionary ecology does today for the macrobiology). Such a theory or family of theories would contain propositions ranging from a few that are applicable to all human societies at all times (e.g., every society is obliged to take account

of variations in the age and sex of its members) to a much larger number that apply only to certain sets and subsets of societies, and even to individual societies. This will not be an easy task. But it is a task we cannot avoid if we accept as our goal the development of a theory capable of explaining all of the more important uniformities and variations in all known human societies, past as well as present—that is the goal set for us by the founders of our discipline in the eighteenth and nineteenth centuries.

A second general desideratum for macrosociological theory is the specification of constants. Many of us learned in graduate school that "you can't explain variables with constants; you can only explain them with other variables." Once again we are confronted with a half-truth. The first half of this proposition is certainly true, at least if you append the qualifier, "alone." But it is also true that you cannot fully explain variables while ignoring relevant constants, and there are always relevant constants in the picture. In regression equations, for example, there is always a regression intercept (i.e., the value of a_1), and the coefficients associated with each of the variables (b_1, b_2, etc.) are also constants.

In macrosociological theory, constants are extremely important. Every society throughout history, for example, has had to deal with certain biological constants (i.e., characteristics shared by all human populations) that interact with social and cultural variables to produce all the various social institutions. As noted above, every society has always had to reckon with the genetic constants of age and sex differences among its members. These have interacted with differences in technology and social organization to produce the various forms of the family that have been observed as well as the various forms of the division of labor between the sexes. Some (e.g., Gordon 1978; Lenski 1966; Lopreato 1984; van den Berghe 1975) have argued that the biological constants to which societies must adapt include much more than the age and sex differences within populations; they may even include the ultimate sources of both cooperation and conflict.[11]

---

[11] Marx's utopian vision of a communist future, in which freedom, justice, and equality would prevail and in which individuals would respond to moral incentives and no longer require material incentives, provides a good

Though certainly debatable, these are issues that should not be swept under the rug as they too often are. Every macrosociolo-. gical theory should contain an explicit statement of assumptions about the nature of human nature (i.e., the relevant effects of our species' common genetic heritage on human action) and other constants. While these cannot be tested *directly* at present (and, perhaps, never will be), models incorporating alternative assumptions *can* be tested and the goodness of fit of relevant data used to draw inferences about the validity of the various options. The present state of affairs, in which assumptions about human nature are usually left implicit (and often, one suspects, below the level of consciousness of theorists themselves) is hardly satisfactory.

We must also consider incorporating *environmental* constants into our theory. These constants, like biological constants, interact with social and cultural variables to produce many striking patterns in human life. To cite but a single example, one cannot help but note the peculiar geographical patterning of societal economic development in the world today. The most advanced industrial societies (as well as those that now appear to be making successfully the transition to an advanced state of industrialism) are almost all in the temperate zones, while the economically most depressed Third World societies are in the tropics. One would suppose that this would suggest an interaction between an environmental constant and technological variables as one of the more important causes of contemporary differences in societal development, but one looks in vain for recognition of this in currently fashionable world-system and dependency theories.[12]

One consequence of this neglect of biological and environmental constants is the theoretical isolation of macrosociological theory in particular, and social science theory in general, from theories in the natural sciences, as noted earlier in this paper.[13] By failing to incorporate relevant assumptions about human nature and the environment that are grounded in current work in the natural sciences—especially biology and geology—we have isolated ourselves and our discipline unnecessarily from the larger scientific community. While this aspect of our situation has seldom been noted or commented on, I suspect that it has a significant effect, at a subliminal level, on the thinking of scientists and the general public alike.

SUMMING UP

Macrosociologists have been complacent for much too long about the state and quality of their theory. For too long we have hidden behind the comforting notion that, because humans are so different from the objects studied in the natural sciences, social scientists are entitled to ignore standards of theory formulation that have evolved over the years in the natural sciences. We have told ourselves that science is merely a method of acquiring knowledge. But science is more: it is also a rigorous and highly disciplined mode of reasoning about causal relationships. This aspect of science cannot be ignored by those who desire the respect and other rewards that modern societies accord the sciences and their practitioners.

It may be too much to say that macrosociology faces a crisis because of the inadequacies of its theories. We have limped along for decades, and we can probably continue.

---

illustration of how implicit assumptions about human nature can create. major problems for macrosociological theory. Marx's failings in this regard were far more excusable, however, than comparable failings of contemporary theorists, since sciences such as genetics and comparative primatology did not exist when he was writing and it was still reasonable to assume that nurture was everything and nature nothing. Moreover, there was no record in his day of the many harmful, and sometimes disastrous, consequences of large-scale utopian experiments (e.g., as in Soviet, Chinese, or Cuban society) on which Marx could base his judgments.

[12] One might suppose that shifts in the locus of societal development over millennia undermines the argument for an environmental constant. This is not the case, however, for those that have occurred during the last 10,000 years, or since the end of the last Ice Age. Although it is true

that more of the most advanced societies were once found in lower latitudes than is the case today (e.g., Mesopotamia and the circum-Mediterranean area), the change is explainable in terms of changes in technology, such as the invention of the heavy plow. This is, of course, only what one should expect if societal development is the result of the *interaction* of technological variables with an environmental constant.

[13] Human ecology and biosociology (not sociobiology) are two notable exceptions to the sociological tendency to ignore the biological bases of human life. Both have sought to build bridges to the biological sciences in a nonreductionist manner. Unfortunately, both efforts have often been received with a mixture of indifference or abuse.

limping for decades to come. But if we allow this to happen, we can expect the number of theories in our field to multiply to the point where knowledgeable observers begin to compare our efforts to the Tower of Babel.

Fortunately, the changes needed to avoid this are not as drastic as they may at first appear. What is chiefly required is more thought and less verbiage. Such thought, however, must lead to more rigorous formulations of theory than those to which we have grown accustomed.

Such a development, working together with the more rigorous modes of data collection and analysis that have become possible in recent years, could easily lead to striking advances in macrosociology. This is the alternative that challenges us, the opportunity that lies before us. This opportunity will never become a reality, however, without acknowledgement of the essential unity of science and of the demands that this unity places on macrosociological theory and theorists.

REFERENCES

van den Berghe, Pierre. 1975. *Man in Society: A Biosocial View*. New York: Elsevier.

Black, Max. 1961. *The Social Theories of Talcott Parsons: A Critical Examination*. Englewood Cliffs, NJ: Prentice-Hall.

Blau, Peter and Otis Dudley Duncan. 1967. *The American Occupational Structure*. New York: Free Press.

Chafetz, Janet Saltzman. 1984. *Sex and Advantage: A Comparative Macro-Structural Theory of Sex Stratification*. Totowa, NJ: Rowman and Allanheld.

Giddens, Anthony. 1973. *The Class Structure of Advanced Societies*. London: Hutchinson.

———. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

Gordon, Milton. 1978. *Human Nature, Class, and Ethnicity*. New York: Oxford University Press.

Harris, Marvin. 1968. *The Rise of Anthropological Theory*. New York: Crowell.

Lenski, Gerhard. 1966. *Power and Privilege: A Theory of Social Stratification*. New York: McGraw-Hill.

———. Unpublished. *Evolutionary Ecology: Principles and Applications*.

Lenski, Gerhard and Jean Lenski. 1987. *Human Societies: An Introduction to Macrosociology*, 5th ed. New York: McGraw-Hill.

Lopreato, Joseph. 1984. *Human Nature and Biocultural Evolution*. London: Allen & Unwin.

Parsons, Talcott. 1953. "A Revised Analytical Approach to the Theory of Social Stratification." Pp. 92–128 in *Class, Status, and Power: A Reader in Social Stratification*, edited by Reinhard Bendix and Seymour Martin Lipset. Glencoe, IL: Free Press.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Snow, Charles Percy. 1959. *The Two Cultures and the Scientific Revolution*. Cambridge: Cambridge University Press.

Turner, Jonathan. 1984. *Societal Stratification: A Theoretical Analysis*. New York: Columbia University Press.

Weingartner, Rudolph. 1967. "Historical Explanation." Pp. 7–12 in *Encyclopedia of Philosophy*, vol. 4. New York: Macmillan.

Woodbury, Richard B. 1968. "Archaeology: The Field." Pp. 378–86 in *International Encyclopedia of the Social Sciences*, vol. I. New York: Macmillan and Free Press.

CORPORATE POLITICAL GROUPINGS: DOES IDEOLOGY UNIFY BUSINESS POLITICAL BEHAVIOR?*

ALAN NEUSTADTL
University of Maryland

DAN CLAWSON
University of Massachusetts

Pluralists argue that corporations with different economic interests and market orientations are incapable of collective political strategy and actions, while class theorists argue that corporations have sufficient class interests to evolve a collective political class strategy. We examined this controversy using a variant of clique detection methods for business Political Action Committee (PAC) contributions to Congressional candidates in the 1980 election. One large group of corporations emerged based on shared conservative beliefs. No other groups nearly as large were detected. The findings were strong and robust for different levels of political similarity. The identified groups differed from each other economically in various ways, but these differences were not nearly as marked as the political differences. Overall, the evidence lends stronger support to the class rather than the pluralist perspective.

INTRODUCTION

One of the central theoretical issues in analyses of business political behavior is the degree to which business is able to unify and act together as an effective political block. Pluralists argue that elites in general and businesses in particular are unable to achieve a general unity; a group may unite on a particular problem or issue, but for other issues, other groups will form. There may be groups among the elite, but the groups are constantly shifting, with alliances formed and dissolved for each new problem. Therefore, there is not *one group* capable of acting with unity and power on a range of issues. Defense contractors might unite around the defense budget and multinationals on Third World debt, but these are specific and pressing economic interests; corporations are unable to unite around a common ideology. Oil and auto companies might unite to promote highway construction, but would divide on policies designed to raise the price of oil. Each corporation pursues its own interest in both the economic and political marketplaces,

and the interests of each are opposed to, or at least different from, those of other corporations. Because they compete in the economic market, pluralists argue, corporations are unable to unite for political action.

Thus, Maitland states, "In view of the extreme heterogeneity of the business community, it may be questioned whether business has any collective interests at all" (1983, p. 3). Berg and Zald agree, arguing that "businessmen are decreasingly a coherent and self-sufficient autonomous elite; increasingly, business leaders are differentiated by their heterogeneous interests and find it difficult to weld themselves into a solidified group" (1978, p. 137). As Useem (1984) notes:

> . . . if the principle of individual corporate focus in business politics prevailed, each major corporation would express, through its leadership, its unique vision of what the government and nonprofit institutions should do on behalf of business and a cacophony of demands, often contradictory, would be heard coming out of the corporate community. (pp. 57–58)

In this most extreme version, each corporation follows its own logic, and there are no consistent political blocks within the business community.

Many other theorists believe that groups of corporations do cohere and engage in similar political behavior. The issue for them is, what groups will cohere? What will be the basis of unity? Will groups oppose one another? For example, Michael Malbin of the American Enterprise Institute has argued that for campaign contributions, commercial banks form one group while savings and loan banks

form another: "To the extent that the groups are buying anything . . . they seem to be cancelling each other out" (1980a, p. 175). While such views are most often associated with pluralists, many Marxists and others on the left analyze the existence of class fractions and divisions within business based on a wide range of factors.[1]

POSSIBLE BASES FOR CORPORATE GROUPINGS

Do corporations act on separate and unique political agendas? Is part or all of the business community able to weld itself "into a solidified group?" (Berg and Zald 1978, p. 137). To the extent that there are any political groupings within the business community, what is the basis of their unity? Is there a multiplicity of opposing groups, or is there a single dominant group? This paper uses data on corporate Political Action Committee donations to candidates in the 1980 Congressional elections to investigate these questions. The 1980 election is particularly important because it had many of the characteristics of a "critical" or "realigning" election. Such elections are marked by changes in voter alignments, elite groupings, and policy outcomes. While most analyses have focused on voter alignments, it is extremely important to analyze and understand the character of elite groupings at the time of actual or potential critical elections (Burnham 1970, 1982; Ferguson and Rogers 1981, 1986).

We can specify a number of potential bases for political groupings within business; each of these has been proposed by one or more analysts. The following list is only illustrative; many additional factors could be specified, but these are the ones most frequently cited. There could be opposing groups, overlapping groups, or scattered individual corporations unable to form any group, each pursuing a separate political strategy.

1. *Geography or region.* For example, businesses in North Carolina might cohere into a political grouping; this could be based either on a similarity of interests (they all need a healthy economy in the area) or on social-interaction patterns (they play golf together and belong to the same clubs). Studies of the patterns of board of directors' interlock networks have shown strong regional subgroupings (Bearden and Mintz 1984). A number of analysts refer to some sort of Sunbelt/Frostbelt split (Burris 1987b; Davis 1986; Dye 1986; Oglesby 1977; Phillips 1982; Sale 1976), and case studies indicate a high degree of interaction among business executives in local areas (Domhoff 1978; Hunter 1980; Whitt 1982).

2. *Industry.* Members of the oil or defense industries might tend to give to the same candidates. One significant influence here is probably a candidate's committee memberships. Similarities of political behavior by industries could be based either on similarity of interest and political unity around these interests (e.g., all oil companies favor the oil depletion allowance and other special tax benefits for the oil industry) or on a competition for influence with key legislators (e.g., all defense contractors want access to key members of the Armed Services committee, each to argue their product is better than that of the other contractors)[2] (Ippolito and Walker 1980, p. 282; Andres 1985; Eismeier and Pollock 1986; Mizruchi and Koenig 1986; Salamon and Siegfried 1977).

3. *Other economic characteristics.* A large variety of other economic characteristics might be bases for unity, either in all circumstances or at particular political conjunctures. Depending on the current issue, firms might engage in similar political behavior because they all want to promote the interests of labor-intensive business (Ferguson 1984), protectionist or free-trade legislation (Ferguson 1983; Klare 1981; Frieden 1980), more money for education to create a high-tech workforce (Levinson 1987), special tax breaks for highly profitable firms, or a host of other factors. Firms or individuals might develop a common perspective because they are highly interlocked and thus consider issues from several viewpoints rather than in terms of a single company's interest (Bonacich and Roy 1987; Burt 1983; Mintz and Schwartz 1985; Mizruchi 1982; Useem 1982, 1983, 1984).

---

[1] See for example Poulantzas 1974, 1978; Ferguson 1984; Ferguson and Rogers 1981, 1986; Klare 1981; Sale 1976; Oglesby 1977. For empirical tests of these claims, see Burris 1987b; Clawson and Neustadtl 1987a; Clawson, Kaufman, and Neustadtl 1985.

[2] As Mizruchi has argued, whether the similarity of political behavior is a result of cooperation or competition, in either case it produces similar political behavior, and this will have a potentially significant impact on the political action of others (Mizruchi 1986).

4. *Pragmatism.* Past analyses of business political behavior have identified two key political strategies, pragmatic and ideological giving (Handler and Mulkern 1982; Clawson, Neustadtl, and Bearden 1986; Clawson and Neustadtl 1987a). Firms using pragmatic strategies seek access to key legislators to influence the details and wording of bills. They will support any legislator who is powerful and willing to allow them to present their case. Of the 243 PACs in our sample, 85 gave more than 70 percent of their money to incumbents, and 55 gave more than 80 percent. Since there are only 37 committee chairs, and perhaps 100 other key legislators (subcommittee chairs, ranking minority members), and since the average corporation in our sample makes 87 contributions, there is an obvious basis for similarity of political behavior.

5. *Ideology.* Firms that try to influence the ideological character of Congress contribute to key races whose outcome is in doubt and that involve candidates with significantly different political positions. In 1980, to promote conservatism, it was usually necessary to defeat incumbents, so these corporations targeted money to challengers. Since there were relatively few conservative challengers running against vulnerable incumbents, and relatively few conservative incumbents whose re-election was in doubt, there was a plausible basis for unity.

We attempted to check for the influence of a variety of economic factors (including most of those listed in 1, 2, and 3 above) on corporate political behavior. Collectively, these factors can explain about 20 to 25 percent of the variation in corporate political preferences (Burris 1987b; Clawson and Neustadtl 1987a), but these methods do not enable us to determine the existence or strength of possible groupings. For example, a group of corporations might all be pragmatic, targeting their giving almost exclusively to incumbents, but they could be giving to different sets of incumbents.

Network methods have the potential to determine (1) whether there are any groups in the data; (2) how strong these groups are, if their members are tightly linked to one another, and if the group is separate and distinct from other corporations. Once meaningful groups are found we can then (3) test the various hypotheses concerning which

factors determine the formation of corporate political blocks.

DATA

The best available data for testing theories of corporate political division and unity are Federal Election Commission (FEC 1981) records of corporate Political Action Committee (PAC) donations. Our focus is on Congressional elections, because the large number of Congressional candidates makes possible a wide range of corporate PAC strategies. Data on PAC contributions are reasonably complete and accurate, and the behavior of the PAC is an indication of the political preferences of the corporation, not simply of individual donors (Clawson, Neustadtl, and Bearden 1986). For a corporation to be included in this study, it must have contributed at least $25,000 to Congressional candidates in the 1980 election. We examined only contributions to major-party candidates who were in the general election.[3] The 243 corporate PACs selected accounted for approximately 73 percent of all corporate PAC gifts in that election. Because a corporation may have more than one PAC, most of the analyses in this paper are based on 230 corporations.[4]

For each corporation included, a variety of information in addition to the FEC data was collected to facilitate analysis of the political and economic characteristics of the corporations. This information was taken from Moody's, the Security and Exchange Commission 10-K Reports, and individual company annual reports for the year 1980.

CLIQUE ANALYSIS

In our attempt to detect political groups among corporations, we want a method that

---

[3] Major-party candidates in the general election account for more than 90 percent of total corporate donations; almost no money is contributed to minor party candidates, and only 7 percent of all money goes to candidates who lose in the primaries. Limitation to major-party candidates in the general election greatly simplifies later analyses.

[4] There are actually only 224 corporations, but we have treated the various AT&T PACs as separate entities, for two reasons. First, the FEC apparently treats them separately for legal purposes (or at least it reports them separately in summary data on the organizations with the largest PAC contributions). Second, in a follow-up article, we intend to analyze subsequent years; to have comparable data for the post-divestiture period, we maintained the many AT&T PACs as separate entities.

provides for the possibility that (1) there could be no groups of any kind, one group, or many groups; (2) a corporation could be in several different groupings—for example, one grouping based on industry, another based on ideology, a third based on region—or in no grouping at all; and (3) the groups could be generally similar (i.e., four regional groups) or opposed (pragmatics versus ideologicals). Given these particular theoretical and substantive demands, clique detection is more appropriate than structural equivalence analysis because, with clique detection, a corporation could simultaneously be in several different cliques, either overlapping or opposed. In structural equivalence analyses, each corporation is constrained to be in only one group. Clique detection is ideally suited for the discovery of a pluralist reality; its results can place each corporation in a variety of different groups.

A clique is a group in which every member is connected to every other member, and there is no other actor that could be added who would be connected to every member of the clique. More technically, in graph theory a clique is a maximal complete subgraph of a total network of actors. Complete means that lines exist between each pair of points of the clique, or that the distance between each pair of points is one. Maximal means that the addition of any point to the subgraph would cause it to *lose its completeness;* not all points are directly reachable.

It is important to differentiate here between the technical mathematical definition of graph theory, and the way the term clique has generally been used by sociologists. In sociology, the term has been used primarily by sociometricians and refers to a face-to-face group with some sort of strong personal ties between the members.[5] There is nothing in the mathematics of graph theory that requires the line connecting two points be based on either a face-to-face connection or a "maximum strength" connection. *Any* criterion could be used to define what constitutes a line connecting two points. Our procedure

returns to the mathematical conception of a clique in graph theory and uses a criterion not previously used in sociological analyses of cliques, a criterion similar to that used in structural-equivalence analyses.

Our method begins with a corporation-by-candidate matrix, where cell entries equal the dollar amount each corporation contributed to each candidate. The donations of each corporation are then correlated with those of all other corporations, thus constructing a corporation-by-corporation matrix of correlation coefficients. Each element in this "similarity" matrix is a measure of the degree of political similarity between two corporations. After the calculation of the similarity or criterion matrix, various levels of degree of similarity are used as cutoffs in the identification of cliques. Corporate political strategies would be totally similar only if a perfect 1.0 correlation existed between them, but this is too strict a standard. Therefore, the analysis was repeated at three degrees of similarity—correlation levels of 0.4, 0.5, and 0.6. A modified adjacency matrix was produced for each level of correlation. For instance, at the 0.4 level, a correlation of .4 or greater was taken to indicate a line connecting two points. Therefore, each entry in the criterion matrix was set to zero if the correlation was less than 0.4, and set to one if it was equal or greater than 0.4. This matrix was then used for the clique analysis.[6] This procedure was repeated at each level of similarity.

The three levels of similarity might produce generally similar results, but they also might produce markedly different patterns. For example, at the .6 level, we might find several regionally based cliques (all California corporations, all Texas corporations, etc.) of size 3 or 4. At the .5 level those cliques would still exist, but might not have added any members, and could appear relatively insignificant compared to several industry-based cliques (all defense contractors, all oil companies, all food producers, etc.) of size 8 or 9. At the .4 level, the cliques from .5 and .6 levels would still exist, but might not have added any members, and could appear much less important than two cliques of size 15 or 16, one of pragmatic corporations and one of ideologicals.

We have used the level of the correlation

[5] This is so much the case that in Burt's discussion of groupings among sociology methodologists, he has written that "a clique is a 'maximal complete subgraph,' or component, so that experts A, B, and C define a clique if they are connected by mutual, *maximum strength,* relations and no further expert can be added to the clique without loosing [sic] this property of mutuality" (Burt 1978 p. 195).

[6] Corporations correlate with themselves at 1.0; by definition these were set to zero.

coefficient, not face-to-face interaction, as our criterion for a linkage between two actors. This differs from most clique[7] detections and is more similar to structural equivalence, so that we need to consider a debate prominent in much of that work. In structural-equivalence analyses, correlation is one of the two main measures used, and has its share of proponents (Breiger, Boorman, and Arabie 1975; Faust and Romney 1985; Cronbach and Gleser 1953). Others advocate the use of Euclidean distance as a measure of similarity (Burt 1986a; Burt and Bittner 1981; Laumann and Guttman 1966; Laumann, Marsden, and Galaskiewicz 1977; Levine 1972), and we could have used it instead of correlation coefficients. Proponents often debate which is the appropriate measure, and the two approaches can provide markedly different results.

We agree with Burt that whether to use correlation or distance as a measure of network similarity ". . . is an empirical question open to debate on a case by case basis" (1986a, p. 205). The appropriate measure depends on what questions the researcher wants to address. We are asking, given the universe of corporate electoral behavior, do groups of corporations follow similar political strategies?

Does the total amount a corporation contributed matter in the construction of a network of political similarity? It matters a great deal when using distance measures and not at all when using correlation. Using a distance measure means that the effects of the total size of any corporation's contributions play an important role in determining the distance between it and all other corporations. Correlations, on the other hand, remove the impact of this influence (Faust and Romney 1985). If we were interested in the influence of business on election outcomes, the crucial variable would be how much money corpora-

tions contribute, and distance would be the appropriate measure of similarity to use in constructing a matrix of criterion linkages. However, we are interested in differences in basic political strategy. For the purposes of this study, the network methods chosen must be able to detect cliques of corporations regardless of the impact a corporation has on the electoral process. Whether a corporation's total donations are large or small should not influence its probability of being grouped with other corporations that follow similar strategies. This contrasts with a research strategy used to group corporations based on the similarity of their influence on election outcomes. By allowing the size of individual contributions to influence the similarity measure, we diminish the ability to find groupings containing corporations with similar strategies but different total resources. Two corporations, for instance, may differ only in their ability to raise resources for their respective PACs—maintaining the same general political outlook—and fail to be considered linked. The difference might result from the fact that one corporation had sales ten times as large as another corporation and therefore gave ten times as much. For this analysis we are less concerned with the magnitude or impact of a corporation's political donations than we are with the similarity of their political strategies.[8]

---

[7] Cliques are generally thought of as small groups of people that relate much more to each other than to those outside the clique. Objection might therefore be made to our use of the term clique. But if the term can be stretched to include large as well as small groups, corporations as well as people, and relations based on correlations in behavior rather than direct communication, then the use of the term clique is appropriate. We will show that members of the cliques and subgraphs we analyze have many relations with each other and few relations with corporations outside the group. Moreover, use of the term clique provides a useful anchor in relation to the mathematics of graph theory.

[8] There is a potential problem in our corporation-by-corporation correlation matrix; we can see it if we look closely at how the entries in the similarity matrix are calculated. On average, a corporation in our sample made contributions to only 14 percent of the possible candidates, so that 86 percent of the candidates did not receive any money from a given corporation, and we assigned a value of zero. Therefore, when correlating corporate political behavior, the resulting coefficients are positively skewed due to the preponderance of zero donations associated with any two corporations. Imagine two corporations that each gave a total of 10 donations, but gave them to different candidates. These corporations will appear more related due to the 639 times their political behavior varies together (because neither contributed to the candidate) than if we only took into account the nonzero entries. In the long run, it is less important that the coefficients are skewed than to know whether the positive skew of the coefficients is a scalar increase or distributed unevenly across corporations. Conceptually this is easy to correct. We need to recalculate the correlations using only those candidates who received donations from at least one of the two corporations being correlated.

A similarity matrix using this modified correlation model was calculated. To determine whether this new procedure lowers all correlations evenly, or whether the effect is distributed unevenly across corporations, we

INITIAL CLIQUES AND THE COMBINATION PROCESS

The modified adjacency matrix represents all network relations at a specified level of similarity. To group corporations that exhibit similar political behavior on the basis of the linkages between all actors, we used the clique algorithm in GRADAP, *Graph Definition and Analysis Package,* developed in the Netherlands.[9]

The data in Table 1 reveal that a large number of cliques are detected at each threshold level. The number of points (corporations) remains the same at each level (.4, .5, .6), but the number of lines connecting the

Table 1. Number of Uncombined Subgraphs at Three Levels of Similarity

Subgraph Size	.4 Level	.5 Level	.6 Level
27	10		
26	28		
25	72		
24	81		
23	101		
22	90		
21	128		
20	155		
19	151		
18	184		
17	156		
16	139		
15	115	1	
14	74	6	
13	105	9	
12	42	17	
11	47	4	
10	53	14	
9	42	16	
8	72	28	
7	61	24	
6	41	26	1
5	43	15	8
4	59	15	11
3	58	33	12
Total number subgraphs	2107	208	32
Number of lines	2600	745	127

took each corporation's scores on the matrix including all zero entries, and correlated them with the corporation's scores on the matrix, omitting candidates to which neither corporation had contributed. This produced a correlation coefficient for each corporation, measuring the extent to which its scores on the two matrices were similar. The mean correlation between a corporation's scores on the two matrices was .80, with a standard deviation of .09—a sufficiently high correlation that it is probable that either approach could be used. However, Knoke and Burt have shown that significantly different findings can result even from two highly correlated measures (1983). As a further check, therefore, the matrix excluding candidates to which neither corporation had contributed was used to replicate the analysis presented later in the paper. Based on the robustness of our findings, we conclude that the positive bias of the correlation coefficient has little effect on the outcome of subsequent analysis. Throughout this paper, we use the unadjusted adjacency matrices (correlations including all zeros, that is, candidates who received no donations from either corporation). The most important reason for this is substantive. We believe this is the correct decision because the zero entry is a meaningful political similarity: both corporations have decided not to contribute to this candidate. The average corporation in our study contributed $56,662. If it had chosen to do so it could have made donations of over $100 to every race. The decision not to do so, but rather to concentrate contributions on a limited number of races, is a political decision.

[9] GRADAP stands for *GRA*ph *D*efinition and *A*nalysis *P*ackage. It is a graph analysis program that provides flexible graph-entry possibilities, graph management, and many variations of graph representation and analysis. It was developed by the Inter-University Project Group of the universities of Amsterdam, Groningen, and Nijmegen. Version 1.1 licensed at the University of Massachusetts currently supports the analysis and management of up to 6,000 points and 60,000 lines. The general correspondence address for GRADAP is: Inter-University Project Group GRADAP, Technical Center FSW, University of Amsterdam, Roetersstraat 15, 1018 WB AMSTERDAM. The phone number is 020-522.2702. GRADAP uses a clique algorithm based on Luce and Perry (1949) with extensions by Alba (1973). The computer subprogram was based on the work of Bron and Kerbosch (1973).

points is larger at lower correlation levels (2,600 at .4, 745 at .5, and only 127 at .6). As expected, the number of cliques and the size of the largest cliques is larger when the number of lines is larger. At the .6 level the largest clique has 6 corporations, at the .5 level 15 corporations, and at the .4 level 27 corporations. There are over 200 cliques at the .5 level and over 2,000 cliques at the .4 level.[10]

What is the meaning of this multiplicity of cliques? At one extreme, every corporation might be in 1 of the 10 cliques of size 27 (at the .4 level) and 8 of the 10 might not contain any members in common. At the other extreme, the cliques could involve varying arrangements of a small number of corporations such that only 30 corporations were

[10] A large clique contains within it several smaller cliques. However, smaller components of larger cliques are *not* included in the totals we report, nor does GRADAP report them separately.

included in the 10 cliques of size 27. We needed a method to combine cliques that are similar to determine the number of groupings that differ from one another.

Alba and Moore (1978, pp. 173–74) provide a method for combining cliques, and a variant of this approach is also used by Sonquist and Koenig (1975). We have used a similar procedure, with some modification. Since one of our goals is to test pluralist theory, a fair procedure must provide every possibility of finding a pluralist reality. The rules used to combine cliques significantly influence the probability that any two cliques will combine and be treated as one. The easier it is for cliques to combine (that is, the less overlap required between two cliques), the more likely it is that they will merge into a single clique, and the less possibility there is of finding what a pluralist would predict: corporations involved in a wide range of different groupings. For two cliques of size 27, if our rule is to combine whenever there is an overlap of one-third or more of the members, only 9 corporations have to be members of both cliques in order to combine. If the rule is one-half or more, 14 corporations must be in both. If the rule is two-thirds or more, 18 corporations must be in both; and if the rule is four-fifths or more, 22 corporations must be in both cliques in order to combine the two. We used an 80-percent rule (four-fifths).

At the next stage of our analysis, all corporations forming cliques at any specified level were combined if 80 percent or more of the members of the smaller of the two cliques (or, of either clique if they were the same size) were also contained within the larger clique. This procedure was performed iteratively until no more aggregation occurred[11] and was performed for each level of similarity. The initial cliques were maximally connected (a line connected each point to every other point); the groups that result from this combination process, which we call subgraphs,[12] are not maximally connected.

Table 2. Number of Subgraphs after Combination at Three Levels of Similarity

Subgraph Size	.4 Level	.5 Level	.6 Level
120	1		
. . .			
. . .			
. . .			
. . .			
. . .			
62		1	
. . .			
. . .			
. . .			
. . .			
10	1	1	1
9	2		
8	2		
7	1	1	1
. . .			
. . .			
4	44	6	10
3	55	32	12
Total	106	41	24

Later, we provide a statistic and an associated probability that a subgraph with given characteristics would be detected by chance alone.

As Table 2 indicates, at the .4 and .5 levels, after combining cliques there is a single large subgraph and no other subgraph remotely as significant. At the .5 level, 62 corporations combine into a single subgraph; the next largest subgraph has only 10 corporations. At the .4 level, the largest subgraph has 120 corporations, and the next largest only 10 corporations. Even at the .4 level, other than the subgraph of 120, there are only 6 subgraphs larger than 4 corporations. Despite an effort to devise rules that would make it difficult for cliques to combine, we do *not* end up with a multiplicity of subgraphs of significant size. Essentially all cliques combine; after combination the large subgraph is six to twelve times the size of any other subgraph. One subgraph is clearly very much larger than the others, though we will also need to investigate the smaller subgraphs.[13] The results are slightly

_____

[11] We used a program written by Betty Levine and Alan Neustadtl.

[12] When cliques are combined, the resulting groups are no longer cliques in the mathematical sense that every member is linked to every other member. Generally we will refer to the results of the combination as subgraphs, though at times we may simply call them groups. Cliques are a special sort of subgraph; every clique is a subgraph, but most subgraphs are not cliques.

[13] The subgraphs of size 3 and 4 are all cliques; cliques of this size cannot combine unless they are entirely contained within some subgraph that has emerged from the combination of larger cliques. (In a clique of size 4, if 3 of the 4 members are part of the larger subgraph, this still only involves an overlap of 75 percent, not the 80 percent needed for combination.)

different at the .6 level. Instead of a single large subgraph, there is one of size 10 and one of size 7; they share four members, but this is not enough to combine.

STATISTICAL CHARACTERISTICS AND SIGNIFICANCE OF THE SUBGRAPHS

What are the characteristics of these subgraphs that result from combining cliques? We first examine whether they are meaningful in a statistical sense; then we investigate the underlying basis for unity in the subgraph. For a defined subgraph to be considered significant, it must satisfy two criteria. First, it must be internally coherent; the points in the subgraph must be connected to each other to a degree that indicates the "solidarity" of the subgraph. This is known as "completeness" (Alba 1973). Second, the subgraph must be distinct from other points and subgraphs peripheral to it. Alba (1973) calls this subgraph "centripetality".[14] Ideally, one would hope for subgraphs that would be highly connected between all their member points and highly distinct from surrounding points (that is, have few connections to them).

To determine whether a subgraph is coherent and distinct, two of Alba's measures (1973) are calculated. First, a measure of internal coherence is used to determine the degree to which the subgraph or clique approaches graph theoretical completeness. It is simply the ratio of the number of lines joining subgraph members to the possible number of such lines and is given as

$$\frac{2q}{p(p-1)} \quad [1]$$

where q is the number of observed lines in the subgraph and p is the number of points in the subgraph. This measure has a theoretical

range of zero to one, although zero is an impossible outcome. The higher this value, the greater the internal coherence and connectedness of the subgraph.

The other measure indicates the degree to which the subgraph is connected to points peripheral to it. This is calculated as the ratio of the number of lines joining members and nonmembers to the possible number of such lines. For any subgraph of p points taken from a large graph of P points, the number of possible lines joining members and nonmembers is $p(P-p)$. Then, if the actual number of such lines is q, the measure of centripetality is

$$\frac{q}{p(P-p)} \quad [2]$$

Ideally, this measure, which ranges from 9 to 100 percent, will be small. Together, to define a coherent and distinct subgraph, completeness should be higher than we would expect given the average density of the graph and centripetality should be lower. These two measures are presented in Table 3; members of the large subgraph at each level are clearly much more connected to each other than to corporations that are not members of the large central subgraph. Alba also demonstrates how to determine the probability of finding subgraphs of specified sizes, completeness, and centripetality; all of our results are statistically significant ($p<.001$).

This main subgraph is so large that the degree of internal connection is not high. At both the .4 and .5 level, about 31 or 32 percent of all possible lines between subgraph members appear. The subgraphs are meaningful and distinct, since only about 1 percent of the possible lines between subgraph members

---

[14] Terminology for these characteristics has not been standardized. Alba refers to the first dimension, the number of lines connecting subgraph members as a proportion of all possible lines connecting them, as cohesiveness or completeness; we call it completeness. The second dimension, the number of lines connecting subgraph members to points not in the subgraph as a proportion of all possible such lines, is called by several different names. Alba (1973) calls it the centripetal-centrifugal dimension, then seems to switch to calling it just centripetal. Sonquist and Koenig (1975) call this centrifugal. We have called it centripetality.

Table 3. Statistical Characteristics of Subgraphs (Percent)

	Complete-ness	Centri-petality
Main subgraphs		
.5 level ($N=62$)	31.04	.84
.4 level ($N=120$)	32.35	1.18
.6 level		
10-member subgraph	71.11	2.14
7-member subgraph	90.48	2.75
Minor subgraphs		
.4 level 10-member	71.11	2.73
.5 level 10-member	86.67	8.09

Note: All results significant at the $p<.001$ level. See Alba (1973) for a discussion of the test of significance.

and other corporations actually exist. The limited completeness is largely a product of the density of the graph and the size of the largest subgraph. The lower the density of the entire graph and the greater the size of the largest subgraph, the lower the maximum possible density for that subgraph. For example, Alba and Moore (1978, p. 181) found only one large subgraph ($N=227$); it included about one-quarter of their population but was only 3.8 percent complete. For our data, if every line in the entire graph were contained in the large subgraph, it would still only be 36 percent complete at the .4 level, and 39 percent complete at the .5 levels.

The limited completeness of the subgraphs that results from combining cliques raises a concern that the average level of political similarity between members of the groups is relatively low. Subgraph members might be correlated at the specified level or above with members of the original clique, but might have very low correlations with subgraph members that were in different cliques initially. For validation, we computed the average correlation between all members of each subgraph. We found that the average level of correlation between the 110 corporate PACs that were not members of any group was .10, the average correlation for the 120 members of the .4 level subgraph was .35, and the average correlation for the 62 members of the .5 level subgraph was .45. This supports our procedure and indicates that corporations were effectively grouped into blocks with similar political behavior.

So far we have focused on the main subgraph. The other subgraphs were much smaller, but what were their characteristics? Most of them are primarily composed of members of the large subgraph. For example, at the .5 level, other than the main subgraph of size 62, there is 1 subgraph of size 10, 1 of size 7, 6 of size 4, and 32 of size 3. In the subgraph with 10 corporations, 7 of 10 are also members of the large subgraph; in the subgraph with 7 corporations, 5 of 7 are (i.e., in each case one less than needed to combine using our 80-percent rule). The 5 corporations in these two subgraphs that are not members of the 62-corporation subgraph are all members of the .4 level-subgraph of 120 corporations. The subgraphs with 4 corporations (which are all cliques in the technical sense) at the .5 level all contain 3 members of the 62-member subgraph; the fourth member is

always a member of the .4 level 120-member subgraph. Similarly, at the .6 level, all the corporations in 4 member subgraphs are also in the .5 level 62-corporation subgraph.

At the .4 level the results are somewhat different. There are 6 subgraphs from size 7 to 10. Both of the 8-member subgraphs, and 1 of the 9-member subgraphs, are composed primarily of corporations in the 120-member subgraph (75 percent and 78 percent respectively). A majority of the members of the other 9-corporation subgraph are members of the 120-member subgraph (5 of 9). Two of the subgraphs are independent, however, with only 1 member from the 120-member subgraph, and all the rest nonmembers. These two subgraphs, one of size 10 and one of 7, will have to be analyzed separately at a later point. Each has completeness and centripetality characteristics indicating that it is statistically significant ($p<.001$). They are the only two groupings in the entire data at any of the three levels (.4, .5, .6) that are clearly separate from the major subgraphs.

UNDERLYING BASIS OF THE SUBGRAPH

What is the meaning of the various subgraphs, especially the large ones at the .4 and .5 levels? What is the basis for their coherence? Do the members of the subgraph share a geographic region, an industry, some economic characteristic, a pragmatic emphasis, an ideological orientation?

The clear unifying basis of the large subgraph at both the .4 and .5 levels is conservativism. At each level, large and statistically significant differences appear[15] between the members of the large subgraph and the corporations not in the subgraph on a range of political variables that measure or are associated with conservatism. As Table 4 indicates, members of the main subgraph are more conservative than their nonsubgraph counterparts. Using a measure we developed where lower negative scores indicate greater liberalism and higher positive scores indicate greater conservatism,[16] at the .4 level, the

---

[15] All tests of significance in Tables 4 through 6 employ t-tests.
[16] See Clawson, Neustadtl, and Bearden (1986) for a fuller discussion of this measure. The measure is based on the similarity of a corporate PAC's behavior to that of a group of 13 noncorporate PACs that explicitly define

Table 4. Political Characteristics of Subgraph Members and Nonmembers (Mean for Category)

	Nonmembers	Members
	$N = .4 = 110$	$N = .4 = 120$
	$.5 = 168$	$.5 = 62$
	$.6 = 220$	$.6 = 10$
Conservatism rating		
.4 Level	0.82	3.44***
.5 Level	1.50	4.06***
.6 Level	2.13	4.21
Voting record conservatism measure		
.4 Level	53.3	65.7***
.5 Level	56.8	67.9***
.6 Level	59.6	67.6
Percent to incumbents		
.4 Level	75.0	43.8***
.5 Level	67.0	35.9***
.6 Level	59.5	32.4
Total dollars contributed		
.4 Level	46522.	59176.***
.5 Level	48804.	64792.***
.6 Level	52351.	72456.

* = $p < .05$
** = $p < .01$
*** = $p < .001$

average rating for corporations not in the subgraph was .82, compared to 3.44 for members of the large subgraph. At the .5 level, the conservatism rating is 1.5 for nonmembers and 4.1 for members of the large ($N = 62$) subgraph. The .6 level subgraph is far smaller but is again based on shared conservative political behavior. Subgraph members also gave a much higher proportion of their money to Republicans. A measure of candidate voting record (the weighted average of *National Journal* ratings) shows the same pattern: subgraph members are more conservative than other corporations.

Political factors other than conservatism produce the same results. In 1980, pragmatic corporations gave to incumbents, and ideological ones to challengers, because the only way to change the ideological character of con-

gress was to defeat incumbents. Corporations in the large subgraphs at each level gave only 44, 36, and 32 percent (at the .4, .5 and .6 levels) of their money to incumbents, while nonmembers gave 75, 67, and 59 percent. Subgraph members also gave much less money than nonmembers to the chairs of Congressional committees and subcommittees (not shown in table). Finally, subgraph members were politically more active, contributing an average of at least $12,000 more than nonmembers. In statistical terms, all these differences are highly significant.

Virtually all members of the main subgraph are conservative, and most conservative corporations are members of the main subgraph, but there is no one-to-one correspondence. Ninety-five corporations have scores of 3.0 or more on our conservatism rating, and 47 have scores of 4.0 or more. While almost all members of the subgraph are unequivocally conservative, a few are not. Of the 62 members of the large subgraph at the .5 level, 55 have conservatism scores of 3.0, but 7 have scores of less than 3.0.[17] Similarly, while most conservative corporations are in the subgraph, some are not. Of 47 corporations with scores of 4.0 or more on our conservatism rating, 37 are in the .5 level 62-member subgraph and 10 are not. Nine of the 10 *are* in the .4 level 120-member subgraph, but *the* most conservative corporation in the entire population, Western Company of North America, is not in the subgraph even at that level.[18]

By far the most important finding is that virtually every group in the data set is based on shared conservative political behavior. But at the .4 level we did find two subgraphs that were not primarily composed of members of the large 120-corporation subgraph. What are the characteristics of these two subgraphs, one of size 10 and one of size 7? The 10-member subgraph contains 8 large defense contractors (all with contracts of $300 million

themselves as ideological conservatives. For the candidates, the mean for this measure is .033 and the standard deviation is 3.05. Scores for the corporate PACs are based on the weighted average of the candidates to whom they contribute. A score of 4, for example, indicates that an average dollar contributed by the corporation went to a candidate who had received donations from 4 of these 13 ideologically conservative PACs. Because corporations favor conservative candidates, the mean for our corporate PACs for this measure is 2.20 and the standard deviation is 1.86.

[17] Abbott Labs, 2.4; Alcoa, 2.8; B.F. Goodrich, 2.8; Motorola, 2.9; Pfizer, 1.7; Signal Companies, 2.6; United Technologies, 2.8.

[18] The conservative corporations not in the .5 level clique and their conservatism scores are Akzona, 4.2; Barnett Banks, 4.7; Blue Bell, 5.6; Brunswick, 4.2; Cargill, 4.3; Clark Equipment, 4.7; Conoco, 4.3; Morrison-Knudsen, 4.7; Northwest Industries, 4.8; Western, 6.1. (Western is one of only three corporations with conservatism scores over 6.0. The other two are Coors and Texas Oil.)

or more) and 2 airline companies. For the 9 corporations that are not members of the 120-member subgraph, the highest conservatism rating is only .5, and all give at least 82 percent of their money to incumbents (the one corporation that is also a member of the 120-member subgraph, General Electric, has a conservatism rating of 1.3 and gives 72 percent of its money to incumbents). We have not been able to determine a unifying basis for the 7-member subgraph; all the members are pragmatic (giving from 75 to 90 percent of their money to incumbents) and nonideological (with conservatism ratings of − .4 to .5), but they do not share a region, industry, or other obvious basis of agreement.

We stress the robustness of our analysis. Subgraphs could potentially have been based on region, industry, economic size, multinational orientation, experiences with government regulation, director interlocks, pragmatism, or a host of other factors. The largest subgraphs at one level might have had a different underlying basis than the largest subgraphs at another level. And corporations could have been in a multiplicity of different subgraphs, each with a separate organizing principle. In fact, whether we look at the large .4-level subgraph of 120 corporations, the large .5-subgraph of 62 corporations, or the .6-level subgraphs of 10 or 7, we always find that shared conservativism provides the unifying basis, and that no other factor can easily be identified.[19] As Appendix 1 indicates, these are the same corporations: every member of a .6-level subgraph was in the .5-level subgraph, and every member of the .5-level group of 62 was in the .4-level group of 120. The same underlying basis appears at all levels, with successively smaller groups as we raise the criteria for membership. Not only did all levels of the subgraph analysis produce the same picture, but a structural-equivalence analysis also gave virtually identical results, identifying one large grouping with no others remotely as large, and that group contained

the same corporations that are members of the subgraphs.[20]

INTERPRETATION AND DISCUSSION

Network methods have revealed characteristics of the data that could not have been established on a statistical footing by other approaches. The particular variant of network methods we have developed demonstrates that most of the plausible bases for corporate political groupings were insignificant in the 1980 election. There were over 2,000 cliques at the .4 level, and over 200 cliques at the .5 level, so it would have been possible for groups to form on a wide variety of bases.

But when cliques were combined using a rigorous rule requiring 80 percent overlap in order to join two cliques, the multiplicity of cliques resolved into a single very large subgraph. The *only* remaining subgraphs are less than one-tenth the size of the large subgraph. What at first appeared to be a pluralist world of hundreds of overlapping groups resolved into a single large group. This group is quite cohesive and separate from the other corporations. There are many connections between members of the group and few connections between them and nonmembers.

The unifying basis for this large subgraph is a shared conservatism. Members of the

[19] Virtually the same results, including both the same clique size and the same corporations, were obtained in the analysis using the matrix based on correlations that excluded all candidates who did not receive donations from at least one of the two corporations being correlated (see text above). For the analysis based on this matrix, we used a .3 level of similarity, which roughly corresponds to a .4 level of similarity in the matrix that includes all zero entries.

[20] We used the STRUCTURE program developed by Burt (1986b) for our structural-equivalence analysis. The basic STRUCTURE program can handle a matrix of up to 150 by 150; we altered the source code to enable us to input our entire 230 by 230 correlation matrix. The STRUCTURE program constrains each corporation to be in one and only one group, which makes it less appropriate as a test of pluralism.

Using this program's liberal clustering option, one group emerges as far larger and more tightly connected than any other, just as with the clique detection. With structural equivalence, it is possible to examine the groups that emerge at any specified distance. At a distance of 1.400, there was one large group with 23 corporations; at the 1.476 level there was one large group with 59 corporations. (The next largest group at this level contained 6 corporations, all transportation companies.)

Of the 23 corporations at the 1.400 level, all are in the .5 level 62-member subgraph, 15 are in the .5 level core and 22 are in the .4 level core. Of the 59 corporations at the 1.476 level, all are in the .4 level 120-member subgraph, and 48 are in the .5 level 62-member subgraph.

The structural-equivalence results thus reproduce the findings from the clique detection. We are preparing a separate methodological article comparing the strengths and weaknesses of the various techniques.

subgraph contribute in ways quite similar to conservative ideological PACs—their average donation goes to a candidate who had received donations from three or four conservative ideological PACs. Subgraph members also give to candidates with more conservative voting records. They are far less pragmatic than other corporations, giving about a third of their money to incumbents, while nonmembers give two-thirds of their money to incumbents; they also contribute only about a third as much to the chairs of committees and subcommittees. On average, subgraph members contribute substantially more money than nonmembers.

Given that subgraph members are conservative and nonmembers tend to be pragmatic, are there any economic factors that explain the contribution strategies chosen by corporations? Data in Table 5 show some relationship between economic characteristics and subgraph membership. At both the .4 and .5 levels, subgraph members are significantly less likely than nonmembers to be in a regulated industry; at both levels they also have significantly fewer director interlocks. These are the only statistically significant differences between members and nonmembers of the .5 subgraph. There is a straightforward explanation why regulated industries are less likely to be members of a conservative subgraph. Subgraph members support challengers in order to change the composition of Congress; regulated industries are so heavily dependent on the specifics and details of Congressional legislation that they pay a higher than usual cost for opposing incumbents. Hence, whatever their personal preferences, they are less likely to join others in conservative ideological strategies.[21] For the same reason, major defense contractors find it difficult to pursue a conservative ideological strategy (Clawson and Neustadtl 1987a).

When we look at total corporate sales and foreign sales, the results appear puzzling. At

[21] A corporation that contributed money to the challenger of a powerful incumbent might make an active enemy. At a minimum, the corporation would find it more difficult to gain access to the member of Congress than if they had contributed to the incumbent. This access has its primary impact on the details of wording of bills; minor wording can save (or cost) a particular corporation tens of millions of dollars (Edsall 1984, p. 110). Some corporations have relatively little need for access, but defense contractors and corporations in regulated industries are highly likely to want such access.

See Burris (1987a) for examples of conservative individuals directing pragmatic corporate PACs."

Table 5. Economic Characteristics of Subgraph Members and Nonmembers (Mean for Category)

	Non-members	Members
	$N = .4 = 110$ $.5 = 168$ $.6 = 220$	$N = .4 = 120$ $.5 = 62$ $.6 = 10$
Percent of group in a regulated industry[a]		
.4 Level	32.7	9.2***
.5 Level	25.3	6.5**
.6 Level	20.6	10.0
Defense contracts as percent of sales		
.4 Level	5.59	.97**
.5 Level	3.74	1.50
.6 Level	3.19	1.39
Director interlocks[b]		
.4 Level	8.1	5.5*
.5 Level	7.5	4.6*
.6 Level	6.7	6.2
1980 sales (millions $)		
.4 Level	4,250.	7,271.*
.5 Level	6,213.	4,969.
.6 Level	5,975.	3,584.
Foreign sales (millions $)		
.4 Level	693.	2,429.
.5 Level	1,692.	1,442.
.6 Level	1,661.	808.
Percent of group with headquarters in Sunbelt[c]		
.4 Level	42.3	49.2
.5 Level	45.6	46.8
.6 Level	47.2	20.0

[a] Depending on who is talking, all industries may be considered regulated, or no industries may be considered regulated. We include the following industries (all of which have been subject to industry-specific regulatory bodies for at least 40 years and most of which are socialized in Western Europe) in this category: banks, insurance, transportation service companies (railroads, airlines, etc.), utilities.

[b] Number of interlocks each corporation has with 200 largest nonfinancial and 50 largest financials (ranked by assets). Data are interlocks in 1976 and were provided by Beth Mintz. The data are described and analyzed in Bearden and Mintz (1985) and in Stokman, Ziegler, and Scott (1985). Even though the data are for four years prior to the 1980 election, a study by Mariolis and Jones (1982) comparing 1962 and 1966 interlock data indicates the correlation is probably about .9.

[c] As used here, "Sun Belt" includes five of the Census Bureau's nine regions: South Atlantic, East South Central, West South Central, Mountain, and Pacific. However, we also tried a number of other definitions, and no relationship exists with any of a variety of other alternatives.

$* = p < .05.$
$** = p < .01.$
$*** = p < .001.$

the .5 level, nonmembers have average sales that are more than $1.2 billion larger than subgraph members, but at the .4 level nonmembers have sales that are $3.0 billion

smaller than subgraph members (a statistically significant difference). A similar situation exists for foreign sales (though the difference is not statistically significant). Table 6 presents the means for three groups: the 110 corporations that are not members of the .4-level subgraph, the 58 corporations that are members of the large .4-level subgraph ($N=120$) but not of the .5-level subgraph, and the 62 corporations that are members of the .5-level subgraph.

Most of the differences between the groups in Table 6 are statistically significant. The most interesting data have to do with sales and foreign sales. The middle group, those we could classify as the conservative "periphery" (since they qualify for the .4-level subgraph that includes half of the population, but not for the .5-level subgraph that includes one-quarter of the population), have significantly larger total sales and foreign sales. Their total sales are about $5 billion larger than either of the other two groups, and their foreign sales are $2 billion larger. They also have the highest proportion of their total sales in foreign markets—35.8 percent, compared to 16.3 percent for corporations not in the .4-level subgraph, and 29.0 percent for corporations in the .5-level subgraph. But this periphery includes many of the largest corporations in America.[22] Our population of corporations includes 13 with sales over $15.0 billion, 9 of them are in the conservative periphery. More than one-quarter of the corporations in the conservative periphery

have sales of $10 billion or more (27.6 percent), while only 11.3 percent of the large .5-level subgraph ($N=62$), and less than 5 percent of the corporations not in the large .4-level subgraph ($N=120$) do. While Coors Beer may be the business most identified with the right-wing resurgence, and is in fact a very conservative corporation, funding for conservative candidates, and the conservative impact more generally, are also dependent on a conservative periphery of *very* large companies. This directly contradicts the claims of corporate liberal theorists (Weinstein 1968), but supports Domhoff's contention that "the leanings of the moderates within the power elite usually determine the outcome of any policy struggle" (1987, p. 179; see also Domhoff 1979, pp. 118–19).

Table 6 also indicates there is no relation between being headquartered in the Sunbelt and membership or nonmembership in the various subgraphs. There is a relationship between these three categories and (1) the proportion of corporations in regulated industries; and (2) the number of director interlocks, with the latter just missing qualifying as statistically significant at the $p<.05$ level.

Shared conservative ideology produces a larger degree of unity than any other factor. In 1978 Ivar Berg and Mayer Zald wrote that "businessmen are decreasingly a coherent and self-sufficient autonomous elite; increasingly, business leaders are differentiated by their heterogeneous interests and find it difficult to weld themselves into a solidified group" (1978, p. 137). Three years after the election we have analyzed, Maitland "questioned whether business has any collective interests at all" (1983, p. 3). While it appears that narrow economic factors such as industry, region, and multinational orientation did not provide a basis for unity by significant

---

[22] Atlantic Richfield, $24.2 billion in sales; Cargill, $11.3; Conoco, $18.7; Exxon, $110.4; General Electric, $25.0; General Motors, $57.7; General Telephone, $10.0; K-Mart, $14.2; R.J. Reynolds, $10.4; Sears Roebuck, $25.2; Shell, $19.8; Standard Oil of California, $41.5; Standard Oil of Ohio, $11.3; Texaco, $51.2; Union Carbide, $10.0; U.S. Steel, $12.5.

Table 6. Economic Characteristics of Corporate Groups

	Not Members of .4-level Subgraph ($N=110$)	Members of .4-level Subgraph but not of .5-level ($N=58$)	Members of .5-level Subgraph ($N=62$)	Sig.
Percent of group in a regulated industry	32.7	12.1	6.5	$p=.008$
Defense contracts as percent of sales	5.6	0.4	1.6	$p=.011$
Director interlocks	8.1	6.4	4.6	$p=.054$
1980 sales (millions $)	4,250.0	9,732.0	4,969.0	$p=.008$
Foreign sales (millions $)	693.0	3,484.0	1,442.0	$p=.045$
Percent of group with headquarters in Sunbelt	42.3	51.8	46.8	$p=.515$

blocks of business, corporations emphatically were able to unite into a large but cohesive group sharing a strategy of donating to key conservative candidates.

If conservativism did not bind groups, there would be almost no groups in our data. At the .6 and .5 levels, there are essentially *no* other groups larger than 3.[23] At the .4 level, there are groups of 7 and 10 members that formed on a different basis. The 10-member subgraph contains 8 defense contractors and 2 regulated airlines, all of them quite pragmatic; the 7-subgraph shares a pragmatic contribution pattern, but has no other clear basis of unity. The *lack* of other groups is itself a major finding. One possible explanation is suggested by Grenzke's (forthcoming) report that the 10 PACs she analyzed (unfortunately, none of them corporate), focused on very different issues. Each of her PACs rated incumbents on their voting records, but there was little overlap in the votes they considered important. Of all the votes used by any of the PACs in constructing their ratings, 64 percent were included by only one PAC, and another 23 percent by only two PACs. It appears that when attention is limited to specific pragmatic interests, pluralists are correct: there is not much basis for strong corporate groupings.

Our findings help shed new light on the 1980 election. Most analyses of "critical elections" focus primarily on voting alignments, but the character of elite groupings is equally important. In the 1970s and 1980s, changes in voter preferences and alignments were not impressive (Ferguson and Rogers 1986). Reagan was not especially popular when he carried through his most impressive policy changes (Clawson, Kaufman, and Neustadtl 1985; Clawson and Clawson 1987), but nonetheless the policy changes have been the most far-reaching since the Great Depression (Burnham 1982; Chubb and Paterson 1985; Peschek 1987). One explanation for this is that a very large and cohesive group of corporations acted on an ideologically conservative understanding—a group not matched

by any other corporate grouping of remotely equivalent size or significance.

Since this is the first such study of corporate political groupings, it is impossible to compare the situation over time. Future research needs to examine (1) whether a similar large and cohesive group of conservative corporations has always existed with little change over time; (2) whether this subgraph emerged in 1980 and then disappeared once it had done its work by providing the basis for a key change in government policy; or (3) whether the subgraph emerged in 1980 and has continued to exist (or strengthen) as an enduring foundation for a rightward shift in American politics.

This study focuses exclusively on similarities in political behavior. The data presented here do not allow us to determine to what degree these behavioral similarities resulted from many corporations acting independently on the basis of the same general criteria, or the extent to which they emerged from purposeful coordination. We note two pieces of evidence that indicate the likelihood of coordination. First, in the 1978 election, the Chair of the Business Roundtable circulated material asking corporations to give more money to "pro-business" candidates, and less to the enemies of business (Handler and Mulkern 1982; Malbin 1980b). Second, in a survey we conducted immediately following the 1986 election, 45 percent of all corporations reported they communicated with BIPAC (Business-Industry PAC) at least once a month, and 42 percent communicated that frequently with the U.S. Chamber of Commerce.[24] Both groups are noted for their conservatism, and both suggest candidates they consider particularly worthy of business support (Clawson and Neustadtl 1987b). The factors most commonly cited in the literature as possible bases for corporate groupings—region, industry, and multinational orientation—all emphasize immediate material interests. Each assumes that if corporations act politically, they must act on the basis of their immediate interests. In the long run, such actions may operate against the best interests of business as a whole, but each firm will have a strong incentive to pursue its own advantage (Maitland 1983). At the turn of the

---

[23] The 7-member subgraph at the .5 level contains 5 corporations that are in the 62-member subgraph, and the other two are in the .4 level 120-member subgraph; conservative ideology is clearly its unifying basis. It has a particular slant of interest, however: all seven corporations (Cities Service, Mobil, Shell, Standard of Indiana, Sun, Texaco, and Union Oil) are large (average sales, $27 billion in 1980) oil companies.

[24] The survey was sent to a random sample of 94 of the 418 PACs that contributed $25,000 or more during the 1983–84 election cycle. The response rate was 58 percent.

century, similar arguments were made about workers: restriction of output could be defeated by introducing piecework, which would give each individual a substantial material incentive to ratebust. In the long run, employers might cut piecerates and return incomes to their previous level, but individual workers would be unable to resist the temptation to maximize their incomes (Clawson 1980).

Whether applied to workers in 1880 or corporations in 1980, this argument in effect denies the reality of class. It assumes that social actors always behave as atomistic individuals pursuing their particular interests without regard to social context. This is

occasionally the case. But, frequently, people understand that their class location and class interests require them to consider more than their short-term needs. At least in corporate contributions to Congressional candidates in the 1980 election, corporations were able to transcend their short-term interests. For both workers (Clawson 1980) and capitalists (Mintz and Schwartz 1985, 1986), this class-conscious activity rests on more than ideology; shared material interests and a set of formal and informal organizations *enforce and coerce* behavior in the interests of the class as a whole.

Appendix 1
Subgraph Members in Groups Larger than 4 and Members of the Core

	.4 Level		.5 Level		
	N=120	Other	N=62	Other	.6 Level
Abbott Laboratories	X[a]		X		
Akzona	X				
Alco Standard	X		X		10
Alcoa	X		X		
Allied Corporation	X				
Amax, Inc.	X	9A			
American Can Company	X				
American Cyanamid	X		X	10	
American Family		9B,7			
American Petrofina	X		X		
American Rice, Inc.		8B			
Amoco	X		X	7	
Armco	X		X		
Ashland Oil	X	9B			
Atlantic-Richfield	X				
Avon	X	9B			
B.F. Goodrich	X		X		
Babcock & Wilcox	X				
Barnett	X				
Bechtel Group, Inc.	X		X		
Belden	X		X	10	
Blue Bell Inc.	X				
Boise Cascade	X				
Bristol-Myers	X		X		
Brown & Root, Inc.	X	9A,8A,8B	X		10
Broyhill Furniture	X				
Brunswick Corporation	X				
Burlington Industries	X				
Cargill, Inc.	X				
Champion International		9B			
Chevron	X				
Cities Service Company	X		X	7	
Clark Equipment	X				
Colt Industries		7			
Columbia Gas	X				
Combustion Engineering	X		X		

Appendix 1 *(Continued)*

| | .4 Level | | .5 Level | | |
	$N=120$	Other	$N=62$	Other	.6 Level
Conoco	X				
Cooper Industries	X		X	10	10,7
Coors Beer	X		X		7
Crown Zellerbach	X				
Dart and Kraft	X		X		10,7
Diamond Shamrock	X		X		
Dow Chemical	X		X		
Dresser Industries	X	8B	X		
E.F. Hutton	X		X		10
Eaton Corporation	X		X	10	10,7
Economics Laboratory	X		X		
El Paso Company	X				
Eli Lilly and Company	X				
ENSERCH Corporation	X	9A			
Exxon	X				
Figgie International	X		X		10,7
Flowers	X		X		
Fluor Corporation	X		X		
General Electric	X	10			
General Mills	X				
General Motors	X				
General Tel. & Electric		10			
Georgia Pacific	X		X		
Getty Oil Company	X		X		
Globe Lighting	X		X		
Greyhound	X		X		
Gulf Resources and Chem.	X		X		
Harris Corporation	X		X		
Hercules, Inc.	X				
Holiday Inns	X		X		
Honeywell	X		X		10
Hughes Aircraft		10			
Hunt Energy Corporation	X		X		
International Harvester	X		X		
International Paper	X				
Intl. Minerals & Chem.	X				
J.C. Penny, Inc.		9B,7			
J.G. Boswell Company	X				
K-Mart	X		10		
Kaiser Aluminum & Chem.	X	9A,8A			
Kellwood	X		X		
Kennecott Corporation	X		X		
L.M. Berry	X		X		
Litton Industries, Inc.	X		X	10	
Lockheed		10			
Louisiana-Pacific	X				
Malone & Hyde	X				
MAPCO, Inc.	X		X		
Marriott	X				
McDonald's Corporation	X		10		
McDonnel Douglas		10			
Merk & Company, Inc.	X		X		
Merril Lynch	X				

Appendix 1 *(Continued)*

	.4 Level		.5 Level		
	$N=120$	Other	$N=62$	Other	.6 Level
Mobil Oil	X		X	7	
Morrison-Knudsen	X				
Motorola	X		X		
Mustang Fuels	X		X		
Mutual of Omaha	X	8A			
Nabisco	X				
NL Industries	X	9A,8A,8B			
Noble Affiliates	X		X		
Northrop		10			
Lone Star Steel	X	8B			
Occidental Petroleum		9A,8A			
Olin Corporation	X		X		
Peabody Coal Company	X				
Pennzoil	X	9A,8A,8B			
Pfizer	X		X		
Pillsbury	X		X		10
Pullman		9B,7			
R.J. Reynolds	X	9B			
R.R. Donnelly & Sons	X				
Raymond International	X		X		
Reader's Digest	X		X	10	7
Republic Steel		9B			
Rockwell International		10			
Santa Fe International	X		X		
Schering-Plough	X		X		
Sears Roebuck	X		10		
Shell Oil	X		7		
Smithkline Corporation	X				
Sohio	X				
Square D Company	X		X		10
Stauffer Chemical Co.	X		X		
Sun Company	X		X	7	
Tempo-Tiger		10			
Tenneco	X		X	10	7
Texaco	X				
Texas Eastern	X	9A,8A,8B			
Texas Oil	X		X		
The Signal Companies	X		X		
TRW		10			
U.S. Steel	X				
Union Camp Corporation	X				
Union Carbide	X				
Union Oil	X		X	7	
Union Pacific		9A,7			
United Energy Resources		8A,8B			
United Technologies	X		X		
Western Airlines		10			
Westinghouse Electric		7			
Weyerhauser Company	X				
Williams Companies	X		X		
Winn-Dixies Stores, Inc.	X	9B,7			

[a] Membership on the two large subgraphs is simply designated by an "X" in the appropriate column. Membership in the other groups is designated by the number of corporations in the group and (where appropriate) by a letter suffix.

REFERENCES

Alba, Richard D. 1973. "A Graph Theoretic Definition of a Sociometric Clique." *Journal of Mathematical Sociology* 3:113–26.

Alba, Richard D. and Gwen Moore. 1978. "Elite Social Circles." *Sociological Methods and Research* 7:167–87.

Andres, Gary. 1985. "Business Involvement in Campaign Finance: Factors Influencing the Decision to Form a Corporate PAC." *PS* 18:213–20.

Bearden, James and Beth Mintz. 1984. "Regionality and Integration in the United States Interlock Network." Pp. 234–49 in *Networks of Corporate Power*, edited by Frans Stokman, Rolf Ziegler, and John Scott. Cambridge, England: Polity.

Berg, Ivar and Mayer Zald. 1978. "Business and Society." *Annual Review of Sociology* 4:115–43.

Bonacich, Phillip and William G. Roy. 1987. "Bank Centrality in 1905." Paper presented at the Sunbelt Social Network Conference, Clearwater Beach, FL.

Breiger, Ronald L., Scott A. Boorman, and Phipps Arabie. 1975. "An Algorithm for Clustering Relational Data with Applications to Social Network Analysis and Comparison with Multidimensional Scaling." *Journal of Mathematical Psychology* 12:328–83.

Bron, C. and J. Kerbosch. 1973. "Finding All Cliques of an Undirected Graph." *Communications of the Association for Computing Machinery* 16:575–77.

Burnham, Walter Dean. 1970. *Critical Elections and the Mainsprings of American Politics*. New York: Norton.

———. 1982. *The Current Crisis in American Politics*. New York: Oxford University Press.

Burris, Val. 1987a. "Business Support for the New Right: A Consumer's Guide to the Most Reactionary American Corporations." *Socialist Review* 91:33–64.

———. 1987b. "The Political Partisanship of American Business: A Study of Corporate Political Action Committees." *American Sociological Review* 52:732–44.

Burt, Ronald S. 1976. "Positions in Networks." *Social Forces* 55:93–122.

———. 1978. "Cohesion Versus Structural Equivalence as a Basis for Network Subgroups." *Sociological Methods & Research* 7:189–212.

———. 1983. *Corporate Profits and Cooptation*. New York: Academic Press.

———. 1986a. "A Cautionary Note." *Social Networks* 8:205–11.

———. 1986b. *STRUCTURE* (Version 3.0). Technical Report #TR2. Research Program in Structural Analysis, Center for the Social Sciences, Columbia University, New York, NY 10027. Included with this report is a computer program "STRUCTURE" that provides algorithms for the analysis of positions in a system of actors.

Burt, Ronald S. and W.M. Bittner. 1981. "A Note on Inferences Regarding Network Subgroups." *Social Networks* 3:71–83.

Chubb, John E. and Paul E. Peterson. 1985. *The New Direction in American Politics*. Washington, DC: Brookings Institution.

Clawson, Dan. 1980. *Bureaucracy and the Labor Process: The Transformation of U.S. Industry 1860–1920*. New York: Monthly Review Press.

Clawson, Dan and Mary Ann Clawson. 1987. "Reagan or Business? Foundations of the New Conservatism." Pp. 201–17 in *The Structure of Power in America: The Corporate Elite as a Ruling Class,* edited by Michael Schwartz. New York: Holmes and Meier.

Clawson, Dan, Allen Kaufman, and Alan Neustadtl. 1985. "Corporate PACs for a New Pax Americana." *Insurgent Sociologist* 13:63–77.

Clawson, Dan and Alan Neustadtl. 1987a. "Interlocks and Corporate Conservatism." Amherst, MA: Xerox.

———. 1987b. "Corporate Political Groupings: Behavior versus Self-reported Communication." Paper presented at the 1987 Nag's Head Conference on Interlocking Directorates.

Clawson, Dan, Alan Neustadtl, and James Bearden. 1986. "The Logic of Business Unity: Corporate Contributions to the 1980 Congressional Elections." *American Sociological Review* 51:797–811.

Cronbach, Lee J. and Goldine C. Gleser. 1953. "Assessing Similarity Between Profiles." *Psychological Bulletin* 50:456–73.

Davis, Mike. 1986. *Prisoners of the American Dream*. London: Verso.

Domhoff, William G. 1978. *Who Really Rules: New Haven and Community Power Reexamined*. New Brunswick, NJ: Transaction.

———. 1979. *The Powers That Be: Processes of Ruling Class Domination in America*. New York: Vintage Books.

———. 1987. "The Wagner Act and Theories of the State: A New Analysis Based on Class-Segment Theory." *Political Power and Social Theory* 6:159–85.

Dye, Thomas R. 1986. *Who's Running America? The Conservative Years*. Englewood Cliffs, NJ: Prentice-Hall.

Eismeier, Theodore J. and Philip H. Pollock III. 1986. "Politics and Markets: Corporate Money in American National Elections." *British Journal of Political Science* 16:287–310.

Faust, Katherine and A. Kimball Romney. 1985. "Does Burt's STRUCTURE Find Structure?: A Critique of Burt's Use of Distance as a Measure of Structural Equivalence." *Social Networks* 7:77–103.

Federal Election Commission. 1981. Computer tape of political action committee donations to candidates for federal office.

Ferguson, Thomas. 1983. "Strategies for a New Economy." *The Nation* 237:108–10.

———. 1984. "From Normalcy to New Deal: Industrial Structure, Party Competition, and American Public Policy in the Great Depression." *International Organization* 38:41–94.

Ferguson, Thomas and Joel Rogers. 1981. "The Reagan Victory: Corporate Coalitions in the 1980 Campaigns." Pp. 3–64 in *The Hidden Election*, edited by Thomas Ferguson and Joel Rogers. New York: Pantheon.

———. 1986. *Right Turn: The Decline of the Democrats and the Future of American Politics*. New York: Hill and Wang.

Frieden, Jeff. 1980. "The Trilateral Commission: Economics and Politics in the 1970s." Pp. 61–75 in *Trilateralism: The Trilateral Commission and Elite Planning for World Management*, edited by Holly Sklar. Boston: South End Press.

GRADAP, The Project Group. 1980. *GRADAP: Graph Definition and Analysis Package User's Manual*, volumes 1 and 2. Inter-University Project Group GRADAP: Universities of Amsterdam, Groningen, and Nijmegen, The Netherlands.

Grenzke, Janet. 1988. "Shopping in the Congressional Supermarket: The Currency is Complex." Forthcoming, *American Journal of Political Science*.

Handler, Edward and John R. Mulkern. 1982. *Business in Politics: Campaign Strategies of Corporate Political Action Committees*. Lexington, MA: Lexington Books.

Hunter, Floyd. 1980. *Community Power Succession: Atlanta's Policy-Makers Revisited*. Chapel Hill: University of North Carolina Press.

Ippolito, Dennis S. and Thomas G. Walker. 1980. *Political Parties, Interest Groups, and Public Policy*. Englewood Cliffs, NJ: Prentice-Hall.

Klare, Michael T. 1981. *Beyond the "Vietnam Syndrome."* Washington, DC: Institute for Policy Studies.

Koenig, Thomas, Robert Gogel, and John Sonquist. 1979. "Models of the Significance of Interlocking Corporate Directorates." *American Journal of Economics and Sociology* 38:173–86.

Knoke, David and Ronald S. Burt. 1983. "Prominence." Pp. 195–222 in *Applied Network Analysis*, edited by Ronald S. Burt and Michael J. Minor. Beverly Hills, CA: Sage.

Laumann, Edward O. and Louis Guttman, 1966. "The Relative Associational Contiguity of Occupations in an Urban Setting." *American Sociological Review* 31: 169–78.

Laumann, Edward O. and Peter V. Marsden. 1979. "The Analysis of Oppositional Structures in Political Elites: Identifying Collective Actors." *American Sociological Review* 44:713–32.

Laumann, Edward O., Peter V. Marsden, and Joseph Galaskiewicz. 1977. "Community-Elite Influence Structures: Extension of a Network Approach." *American Journal of Sociology* 83:594–631.

Levine, Joel H. 1972. "The Spheres of Influence." *American Sociological Review* 37:14–27.

Levinson, David. 1987. *High Technology's Impact on Higher Education*. Ph.D. diss., University of Massachusetts at Amherst.

Luce, R.D. and A. Perry. 1949. "A Method of Matrix Analysis of Group Structure." *Psychometrika* 14: 94–116.

Maitland, Ian. 1983. "House Divided: Business Lobbying and the 1981 Budget." *Research in Corporate Social Performance and Policy* 5:1–25.

Malbin, Michael J. 1980. "Of Mountains and Molehills: PACs, Campaigns, and Public Policy." in *Parties, Interest Groups, and Campaign Finance Laws*, edited by Michael J. Malbin. Washington, DC: American Enterprise Institute.

———, ed. 1980b. *Parties, Interest Groups, and Campaign Finance Laws*. Washington, DC: American Enterprise Institute.

Mariolis, Peter and Maria Jones. 1982. "Centrality in Corporate Interlock Networks: Reliability and Stability." *Administrative Science Quarterly* 27:571–84.

Mintz, Beth and Michael Schwartz. 1985. *The Structure of Power of American Business*. Chicago: University of Chicago Press.

———. 1986. "Capital Flows and the Process of Financial Hegemony." *Theory and Society* 15: 77–101.

Mizruchi, Mark S. 1982. *The American Corporate Network 1904–1974*. Beverly Hills, CA: Sage.

———. 1986. "Determinants of Similarity of Political Behavior Among Large American Corporations." Grant Proposal to the National Science Foundation.

Mizruchi, Mark S. and Thomas Koenig. 1986. "Economic Sources of Corporate Political Consensus: An Examination of Interindustry Relations." *American Sociological Review* 51:482–91.

Oglesby, Carl. 1977. *The Yankee and Cowboy War*. New York: Berkeley.

Peschek, Joseph. 1987. *Policy Planning Organizations: Elite Agendas and America's Rightward Turn*. Philadelphia: Temple University Press.

Phillips, Kevin. 1982. *Post-Conservative America: People, Politics and Ideology in a Time of Crisis*. New York: Random.

Poulantzas, Nicos. 1974. *Political Power and Social Classes*. London: New Left Books.

———. 1978. *State, Power, and Socialism*. London: New Left Books.

Salamon, Lester M. and John J. Siegfried. 1977. "Economic Power and Political Influence: The Impact of Industry Structure on Public Policy." *American Political Science Review* 71:1026–43.

Sale, Kirkpatrick. 1976. *Power Shift: The Rise of the Southern Rim and its Challenge to the Eastern Establishment*. New York: Random.

Sonquist, John A. and Thomas Koenig. 1975. "Interlocking Directorates in the Top U.S. Corporations: A Graph Theory Approach." *Insurgent Sociologist* 5: 196–229.

Stokman, Frans N., Rolf Ziegler, and John Scott, eds. 1985. *Networks of Corporate Power: A Comparative Analysis of Ten Countries*. Cambridge, England: Polity.

Useem, Michael. 1982. "Company Versus Classwide Rationality in Corporate Decision Making." *Administrative Science Quarterly* 27:199–226.

———. 1983. "Business and Politics in the United States and United Kingdom: The Origins of Heightened Political Activity of Large Corporations During the 1970s and Early 1980s." *Theory and Society* 12: 281–308.

———. 1984. *The Inner Circle*. New York: Oxford University Press.

Weinstein, James. 1968. *The Corporate Ideal in the Liberal State: 1900–1918*. Boston: Beacon.

Whitt, J. Allen. 1982. *Urban Elites and Mass Transportation*. Princeton, NJ: Princeton University Press.

THE TRANSFORMATION OF THE SOUTHERN RACIAL STATE: CLASS AND RACE DETERMINANTS OF LOCAL-STATE STRUCTURES*

DAVID R. JAMES
Indiana University

Theories of the state are unable to explain intrastate variation in policy implementation because they tend to focus on the highest levels of the state hierarchy and ignore local-state institutions. Race-relations theories ignore the state entirely. This research views the local state as constrained by the superordinate national state and by local class structures. The racial nature of the local state, here termed the racial state, was designed and defended by plantation operators and white farm owners. Institutional features of the racial state conformed to the interests of planters and white farmers. These two classes also impeded the national state's dismantling of the racial state during the 1960s. Racial differences in enfranchisement were closely articulated with the local class structures typical of labor-intensive cotton agriculture rather than with competition and status variables suggested by race-relations theories.

The white race deems itself to be the dominant race in this country. And so it is, in prestige, in achievements, in education, in wealth, and in power . . . . But in the views of the Constitution, in the eye of the law, there is in this country no superior, dominant ruling class of citizens. There is no caste here. Our Constitution is color-blind and neither knows nor tolerates classes among citizens. In respect of civil rights, all citizens are equal before the law.
<div align="right">John Marshall Harlan (1896)</div>

INTRODUCTION

Current theories of the state emphasize the importance of the national state both as an object of struggle by contending political forces and as a determinant and regulator of economic and social processes (e.g., Przeworski 1985; Skocpol 1979; Wright 1978). But theories that focus on the national state cannot explain the enormous historical variation in the local implementation of national

policies. For example, powerful landlords and farm owners shaped the local implementation of the New Deal farm programs (James 1986; Wright 1986). Unemployment compensation and old-age assistance was manipulated to suit local employers (Alston and Ferrie 1985; Piven and Cloward 1971). The desegregation of public schools has been very uneven (Orfield 1969, 1978). Variation in the implementation of these and other policies cannot be explained by exclusive attention to the national level because all policies were intended to be uniformly enforced. In each case, local political institutions played an important role in bending the enforcement of national legislation to comply with dominant local interests.

The variability in the extension of the franchise to white and blacks in the South across periods and localities illustrates this process. Black voter-registration drives were very successful in border states during the 1960s, but met implacable resistance in core southern states (U.S. Commission on Civil Rights 1961; 1968). This research addresses the importance of local-state structures in maintaining social order and protecting local economic and social institutions.[1]

[1] The term "state structure" refers to the interlocking network of formal organizations that are usually legally constituted and that exercise a virtual monopoly of the means of physical violence and coercion within a certain territory. These interlocking structures include the police, the courts, legislative bodies, city councils, county administrative organizations, executive agencies, and the like.

John Marshall Harlan's famous dissent in *Plessy v. Ferguson* (Harlan [1896] 1970) opposed the legal institutionalization of two distinct racial classes of citizens, but the majority decision made white people the "dominant race" as an aspect of state policy. Black disfranchisement was part of a maze of racially discriminatory local- and state-level policies. Racial segregation was legally mandated and enforced in most spheres of public and private life. Whites held virtually all elective and appointive offices. Local public officials such as governors, judges, mayors, juries, sheriffs, police officers, and registrars were white (e.g., Myrdal 1944; U.S. Commission on Civil Rights 1959). Racial distinctions were explicitly incorporated within state structures, both as procedures and practices and as implicit criteria for state officials. This array of institutionalized discriminatory features, which I call the *racial state,* contrasts dramatically with the *liberal democratic state* that guarantees equal citizenship rights (Marx [1843] 1978, p. 33; Bendix 1964, pp. 86–126).

Neither racial nor liberal democratic states exist in pure forms. Real states typically combine features of both, but theoretical distinctions can be clarified. The liberal democratic state confers equal political rights on citizens regardless of their race, income, education, religion, or other class and status distinctions. The racial state uses racial criteria to assign unequal political rights to different races.[2] The liberal democratic state ignores petitions for redress of grievances based on racial status. The racial state metes out different punishments for equivalent crimes according to the race of criminals and victims. The liberal democratic state assigns children to schools without regard to race; the racial state segregates them. The liberal state protects the voting rights of all; the racial state disfranchises blacks.

Marxist theories argue that the state is determined by class structures in important and systematic ways. Three periods of political crisis in the South are analyzed below to clarify the relationship between class structure and the racial state. The first is the period before and during the Civil War that

led to the abolition of slavery, a severe blow to the racial state. The second occurred during the late nineteenth century and led to racial segregation and black disfranchisement, a strengthening of the racial state. The third dismantled the racial state during the 1960s.

A key measure of the strength of the racial state is the difference between black and white voter-registration rates. Southern blacks were completely disfranchised during slavery, but registered in proportions almost equal to whites during and immediately after reconstruction. The racial state nearly eliminated blacks as voters by 1900, but not as completely as during slavery. Black rates have again approached those of whites since the 1960s.

Few question that a racial state was essential for the survival of antebellum slave agriculture. A consensus is emerging that the postbellum southern racial state was also created and maintained by forces organized to defend the class structure of labor-intensive cotton agriculture (e.g., Bloom 1987; Kousser 1974; Myrdal 1944). The claim that the racial state protected the rights of cotton planters and farm owners against those of farm laborers, tenants, and sharecroppers is more controversial (e.g., Cohen 1976; Woodman 1977; Wright 1986).

This new racial state was less efficient than the slave state in maintaining white privileges and enforcing black subordination. Still, black disfranchisement and state-enforced racial segregation created a politically weak and docile black labor force that was unable to demand higher wages effectively. Physical and economic coercion was the modal strategy used by planters to maintain the profitability of cotton-plantation agriculture. Politically impotent blacks could not defend themselves against the authority of white planters. Even as late as the 1950s, the most productive and profitable cotton farms were labor-intensive and were owned and operated by white planters who exploited the labor of black sharecroppers and tenant farmers (LeRay and Crowe 1959; Pederson and Raper 1954). Southern tobacco farm owners also resorted to coercion when confronted with rising labor costs or declining commodity prices (Badger 1980; Shifflett 1982). By contrast, other white employers of black laborers could use capital-intensive strategies to cope with the vagaries of labor and commodity markets. Consequently, white farm owners (especially

---

[2] Restricting the discussion to race ignores state-enforced discrimination on the basis of other status distinctions, such as gender or religion, that distinguish real states from liberal democracies.

cotton and tobacco farmers) who were actual or potential employers of black farm workers and sharecroppers were probably the most determined defenders of the racial state (Paige 1975).

The specific research question is to what extent did rural class relations, especially those associated with labor-intensive cotton and tobacco agriculture, account for the regional variation in the strength of the southern racial state? First, I examine race-relations theories for alternative explanations of black disfranchisement. Second, I propose a theory of the relationship between local-state and class structures that explains variation in the strength of the racial state across the South. Third, I investigate the class bases of the racial state during the first two crisis periods. I identify specific mechanisms that linked local-state structures to the interests of white farmers and plantation owners. Finally, I assess the linkages between class structure and the racial state during the third crisis period by analyzing the determinants of racial differences in voter-registration rates in 1958, 1964, and 1967 in seven core southern states.

RACE-RELATIONS THEORIES AND BLACK DISFRANCHISEMENT

Theories of race relations provide two kinds of explanations for black disfranchisement. *Stratification* theories argue that stratified social structures create *status distinctions* that are important determinants of the attitudes and values of discriminators. Certain white status groups (for example, the more educated) are more tolerant of blacks or have a weaker "taste for discrimination" (Becker 1971) than others and, therefore, discriminate less. *Competition* theories argue that discrimination stems from competition between different racial groups over *scarce resources* such as political power, land, and jobs (e.g. Horowitz 1985; Wilson 1978; Hannan 1979). Consequently, the principal agents of institutionalized racial discrimination are *prejudiced white status groups,* according to stratification theories, or *white groups confronted with competitive pressure from blacks,* according to competition theories.

Stratification theories identify several social and economic determinants of black disfranchisement. For example, higher black median income and education levels were modestly associated with higher black voter-registration rates in 1958, while higher white education levels and greater concentrations of farm workers and tenants were associated with lower black rates (Matthews and Prothro 1963a). Higher black proportions had a *negative* effect on black voter-registration rates before the 1965 Voting Rights Act (VRA) (Matthews and Prothro 1963a), but a *positive* effect afterward (Daniel 1969). Higher black proportions supposedly stimulated the discriminatory attitudes of whites before the VRA, but encouraged black mobilization afterwards by increasing blacks' perception of favorable election results. These accounts assign causal primacy to attitudes and values; state effects are minor or exogenous to the theory.

A classic version of competition theory states that fear of black political domination motivates all whites to discriminate, but fear of economic competition spurs only the white working class. Both fears lead to increasing discrimination by whites as the minority percentage increases (Blalock 1967, p. 30).

Wilson (1978) agrees that white workers are the biggest enemy of black civil rights. Jim Crow segregation and black disfranchisement in the South accompanied the rise of lower-class whites to power after the Civil War and represented the interests of whites in eliminating economic competition from black workers (Wilson 1978, pp. 59, 137). A small black population in the North prevented the northern white working class from imposing state-enforced segregation and disfranchisement, while a much larger black proportion lost the competitive struggle to white workers in the South (pp. 55–60, 82–87). The theory is flawed. It cannot be salvaged by substituting the white rural classes for the industrial working class as the discriminatory agent because the interests of other powerful actors were at stake. White sharecroppers and laborers may have competed with blacks, but white planters and farmers did not. White planters and farm owners had a common interest in depressing the incomes of sharecroppers and agricultural laborers regardless of their race. Given the divergent interests of white southerners, white solidarity probably reflected the domination of white farmers over the rural under-classes of both races rather than competition between lower-class whites and blacks.

Both stratification and competition theories of American race relations ignore theories of

the state and politics. Both assume that potential discriminators will find institutional mechanisms to express their prejudices. By default, they adopt a pluralist perspective that views the state as either a mechanism for aggregating group preferences or a reflection of societal norms and values (Alford and Friedland 1985, p. 43). Race-relations theories ignore the possibility that state structures can independently affect the mobilization of racial groups or the implementation of racial policies.[3]

THE LOCAL STATE IN THE
UNITED STATES

Local-state structures established the "rules of the game" for practicing racial discrimination in the South before the 1960s. The nature of the local state varied as a function of the relative strength of two sets of constraints: those imposed by the superordinate state and those imposed by local class structure.

Constraints of the Superordinate State

The exercise of sovereignty in the United States is partitioned among federal, state, and local-state institutions that are further subdivided horizontally and vertically. This fragmentation of sovereignty provides differential access to state power for various agents and interest groups and makes uneven policy implementation across local jurisdictions likely. Uneven policy implementation is compounded if local states enjoy considerable autonomy as they do in a federal system like the United States. Local autonomy in the U.S. results not just from decentralization and fragmentation, but also from the local state's capacity to control the recruitment of personnel to local offices and power to tax its citizens to cover the costs of local government (Peterson 1981, p. 68; Williams 1980).

The national state has other mechanisms that constrain local-state autonomy besides appropriation of local power to appoint officials and to tax citizens. Direct federal

intervention bypasses the local state altogether. Occupation by federal troops is an extreme example of this mechanism although it is rarely used. The federal court system also poses a threat to local autonomy but only if its interpretation of the Constitution or federal legislation is disruptive of local power relationships. Southern influence in Congress reduces the threat of direct intervention or disruptive legislation (Bensel 1984). Even if federal legislation were potentially disruptive, the requirement to litigate separately each citizen's claim reduces the courts' effectiveness in compelling uniform compliance with national policy.

The local state in the South had a high degree of autonomy from the national state during the first half of the twentieth century because national policy enforcement mechanisms were weak. Direct intervention was not possible during this period, and the federal courts were ineffective in protecting the civil rights of blacks. Most revenues were raised through local and state taxes and almost all personnel recruitment was locally controlled. The U.S. Department of Agriculture began making direct payments to farmers during the 1930s, but these were placed under the control of local committees that were locally elected, locally controlled, and especially sensitive to the needs of the economically powerful (e.g., James 1986; Wright 1986; Fligstein 1981). Federal welfare payments were also adjusted to the needs of local employers and planters (Alston and Ferrie 1985; Piven and Cloward 1971).

Constraints of Local Class Structure

Because federal constraints on local states were weakly enforced, local political forces strongly affected the nature of local-state structures. Four mechanisms ensured that local states were especially attuned to the interests of cotton planters, farmers, and other employers of black workers. First, all local states had to enhance local economic productivity to protect the fiscal base of local government, which is tied to property values through taxes (Peterson 1981, pp. 17–38). Because labor-intensive cotton and tobacco production were the most important economic activities during most of this period, local officials had to promote them.

Second, the political fortunes of local officials were intimately linked to policies

[3] Hannan (1979), Nagel and Olzak (1982), and Nielsen (1985) acknowledge that the state can affect the structure of race relations and are an exception to this criticism. They view the state as an effect of the competition of ethnic groups but with the state as an additional organized participant in the competitive struggle (Hannan 1979, p. 266).

that enhanced the economic fortunes of local economic elites (Peterson 1981; Block 1977). Local officials were vulnerable to the distribution of local political power and were forced to cooperate with local constituencies to be effective and enhance their job security (e.g. Lipsky 1980; Williams 1980). Cotton planters and white farmers were the most powerful constituency in the rural South. Besides being the local opinion leaders, they routinely bought votes, bribed officials, and used economic coercion to punish their enemies (e.g., Myrdal 1944; Davis, Gardner, and Gardner 1941).

Third, many local officials such as the sheriff, the justice of the peace, and minor court officials were paid by fees for each arrest, warrant served, and conviction obtained (e.g., Myrdal 1944, p. 548; Kirby 1987, p. 217; Davis, Gardner, and Gardner 1941, p. 496). The fee system combined with the criminal-surety system (discussed in the next section) to create a community of interests between law-enforcement officials and cotton planters. Law-enforcement officials needed numerous arrests and convictions; cotton planters needed cheap laborers. Because blacks were politically weak, they satisfied both needs. Sheriffs augmented their incomes by arresting blacks for petty crimes; planters who paid the fines of offenders obtained cheap labor services in return for their beneficence. The remaining prisoners labored on the county farms and roads to reduce local government costs (Kirby 1987; Daniel 1972).

Finally, cotton planters, white farmers, and their allies used physical violence to wrest the reins of local government from their domestic enemies. Nineteenth-century southern history is replete with examples of violence and vigilantism. Blacks, Republicans, and Populists were lynched, murdered, tortured, and terrorized until all were eliminated as viable competition to the Democratic party, the champion of planter interests and white supremacy. Violence was a key conservative weapon from the end of the Civil War until after the Great Depression (e.g., Myrdal 1944; Kousser 1974). Even during the 1960s, fear of violence was still a major barrier to the enfranchisement of southern blacks (e.g., St. Angelo and Puryear 1982; Salamon and Van Evera 1973; U.S. Commission on Civil Rights 1959, 1965, 1968).

The local state in the South did not mediate among contending interest groups, but institutionalized racial bias. The theory accounts for the great variation in institutionalized racial bias across periods and localities in the U.S. by identifying two sets of constraints. When the central state's constraints were strong, local class structures had little effect. When the central state's constraints were weak, local class structures had greater effects on the local state.

CLASS BASES AND FEATURES OF THE RACIAL STATE IN SLAVERY AND FREEDOM

The structure of the southern racial state was transformed as a result of fiercely contested political struggles during and immediately after the Civil War, during the late nineteenth century, and again during the 1960s. On each occasion, the scope and strength of racially oppressive, local-state structures were altered as contending political forces attempted to modify the racial selectiveness of the state in their favor.[4]

Class Structure and Features of the Slave State

All observers agree that the state institutions of the antebellum South were biased in favor of the interests of slave owners and against those of black slaves. The pre-Civil War U.S. Constitution left most issues concerning slavery to the states.[5] The states delegated primary authority and responsibility for slave discipline to slave owners, but this authority was reinforced by the state-level judicial and executive systems and defended by state police power. Slave patrols and state militias augmented the authority of slave owners by

---

[4] Political crises provide valuable opportunities to examine the relationship between the state and political forces based in particular class structures because the greatest changes in states occur during those times. Biases built into the state during crises tend to persist and may seem normal and inevitable after the passage of time. Consequently, the class determinants of state structures may be masked during "normal" intercrisis periods, but should be apparent during periods of crisis.

[5] The U.S. Constitution acknowledged the existence of slavery three ways: (1) by adjusting Census counts for Congressional apportionment and levying direct taxes (slaves were counted as three-fifths of a person); (2) by preventing the banning of the slave trade until after 1808; and (3) by providing for the return of fugitive slaves (Meier and Rudwick 1976, p. 54).

apprehending runaways, punishing slaves for offenses committed away from the plantation, and smashing rebellions (Meier and Rudwick 1976, pp. 59–63; Litwak 1980, pp. 27–28; Stampp 1956, pp. 141–236).

The slave state was essential for the continued existence of plantation slavery and the small elite class of slaveowners who reaped the majority of its benefits (3 percent of white families owned over half of the slaves in 1860) (Stampp 1956). The instability of southern political authority during the Civil War demonstrated this intimate connection (Litwak 1980), and the abolition of slavery confirmed it.

Class Structure and Features of the Racial State

The planter class was weakened politically and economically by the abolition of slavery, but not destroyed. Post-Civil War amendments to the Constitution, Congressionally mandated reconstruction, and U.S. Army occupation made it impossible to reimpose the essential features of slavery, but plantation land was never redistributed to the former slaves.[6] If blacks could be forced to work for low wages using the labor-intensive technology inherited from slavery, the cotton-plantation economy could be resurrected. Conservative political forces based in the plantation regions made this their major political goal for the next generation.

A new form of the racial state was in place throughout the South by the end of the nineteenth century. This new racial state had three important features that enforced black subordination: (1) racial segregation; (2) defense of landlord and employer prerogatives against those of sharecroppers and employees; and (3) black disfranchisement. As a result, blacks were particularly vulnerable to physical and economic coercion and exploitation, although never as completely as during slavery.

The U.S. Supreme Court's decision in *Plessy* eliminated the last legal barrier to Jim Crow segregation and removed the threat of federal intervention in defense of the civil rights of blacks. All important areas of public and private life, with the prominent exception

of labor markets, were racially segregated by law during the first two decades of the twentieth century (e.g., Woodward 1974; Myrdal 1944). Southern states and localities confined blacks to inferior school systems, housing areas, public services and facilities, and stamped blacks with a "badge of servitude" (Harlan [1896] 1970).

The second feature of the racial state, the defense of landlord and employer prerogatives, consisted of a complicated network of laws and practices that interfered with the operation of southern labor markets. Employers and landlords were protected by "enticement statutes" that made it a crime to hire a worker already under contract and by "emigrant agent laws" that retarded the exodus of black workers by charging northern labor recruiters prohibitively expensive license fees (Cohen 1976). Crop-lien laws protected the interests of landlords against those of sharecroppers and laborers. Crop-liens were the only source of credit available to sharecroppers and laborers to cover production costs and to provide for their families' sustenance during the long growing season. Sharecropping, financed through crop-liens, became the preferred form of plantation labor organization (Schwartz 1976; Ransom and Sutch 1977; Wiener 1978).

A variety of laws regulated the behavior of laborers, sharecroppers, and tenant farmers. Contract-enforcement statutes were repeatedly devised that made it a criminal, rather than a civil offense, to break a labor contract (Cohen 1976; Daniel 1972). Broadly drawn vagrancy statutes were used to supply plantations with workers during labor shortages (Cohen 1976). Even as late as the 1940s, city police departments were still conducting vagrancy drives to provide cotton planters with plantation workers (Hoffman 1969). Finally, criminal-surety systems allowed employers to pay fines for individuals convicted of minor crimes such as vagrancy and drunkenness. In return, the miscreant was contractually obligated to repay the benefactor, and labor services were the only possible means of repayment (Cohen 1976; Woofter 1936).

Combined with racial segregation and disfranchisement, these laws made blacks vulnerable to economic discrimination. As a result, southern labor markets became more segregated during the early twentieth century. Blacks were confined to the most menial,

---

[6] The 13th Amendment outlawed involuntary servitude, the 14th guaranteed equal protection of the law, and the 15th extended the franchise to the freedmen.

labor-intensive, and poorest-paying jobs in agriculture and industry. They had lower incomes, poorer living standards, and were preferred to whites as plantation laborers.

Black disfranchisement was the third principal feature of the racial state. Because of the post-Civil War amendments and the threat of federal intervention, the mechanisms of black disfranchisement had to be "color blind" in form but color sensitive in substance. The most effective mechanisms were poll taxes, difficult registration requirements, property requirements, literacy tests, and restricting primary elections to whites (e.g., Kousser 1974; Hine 1979). Many poor and illiterate whites were also disfranchised by these devices, but blacks were eliminated as a political force.

Black disfranchisement was the key to the stability and effectiveness of the racial state (Woodward 1974, p. 83; Key 1949). Because blacks were unable to vote, they could not sanction state officials through routine political processes. Consequently, segregation statutes were administered in a separate and *unequal* fashion. White landlords and employers could discriminate against black sharecroppers and laborers with impunity. Whites were almost never punished for even the most savage physical violence against blacks (Davis, Gardner, and Gardner 1941; Myrdal 1944).

Just as cotton planters provided a determined defense of the slave state, the plantation regions provided the strongest support for the racial state in the late nineteenth century. Planters supplied the leadership and theory for black disfranchisement in almost every case and always gave it their political support (Kousser 1974). Eliminating blacks from the electoral process also eliminated party competition, ensured that lower-class interests would not be effectively represented in political decision making, and protected the interests of "Democrats [who were] usually from the black belt and always socioeconomically privileged" (Kousser 1974, p. 238).

Yeoman farmers in the white-majority, non-plantation counties had mixed motives for supporting the racial state. They too had an interest in subordinating blacks because they were occasional or potential employers of black workers. Black disfranchisement might also reduce the political power of cotton-belt whites who controlled the ballots of large numbers of black voters through fraud and intimidation. But the constitution-

ally permissible disfranchisement devices were very imprecise. "Color-blind" tests such as poll taxes and literacy tests were designed for blacks, but were used to disfranchise many white agrarian radicals and Populists who threatened the hegemony of cotton-belt Democrats. White farmers from the nonplantation regions supported black disfranchisement, but less zealously than cotton planters did (e.g., Kousser 1974; Key 1949; Hahn 1983).

Southern cities provided the weakest support for the racial state. Southern industrialists, less dependent on labor-intensive technologies, were less likely than cotton planters and farmers to use coercive methods of labor control (Paige 1975). Plantation owners preferred black laborers, but industrialists were content to hire whites. Nevertheless, cities, containing only 20 percent of the southern population in 1900, were economically and socially linked to the rural areas. Rural demagogues found urban allies to impose racial segregation and black disfranchisement in the cities (Rabinowitz 1978). Blacks were effectively barred from industrial employment, especially in the higher-paying jobs, and unions were segregated (Northrup, Rowan, Barnum, and Howard 1970; Wright 1986). Urban imposition of Jim Crow segregation and black disfranchisement consummated the regional victory of plantation racial politics.

The Persistence and Decline of the Racial State

The racial state denied blacks civil rights for more than 50 years through depression, war, rising urbanization and industrialization, and the declining economic importance of cotton and tobacco farming. Capitalist economic development severed the link between the economic fortunes of most white southerners and the subordination of black agricultural workers (Bloom 1987; Mandle 1978). Yet, the racial state continued to reward bigotry (Barkan 1984) and make discrimination appear endless and natural in the everyday affairs of whites.

Black and white voter-registration rate trends reflect the strength of the racial state during the first two-thirds of this century (see Figure 1). Louisiana was the only state that routinely collected registration statistics by race for the early part of the century.
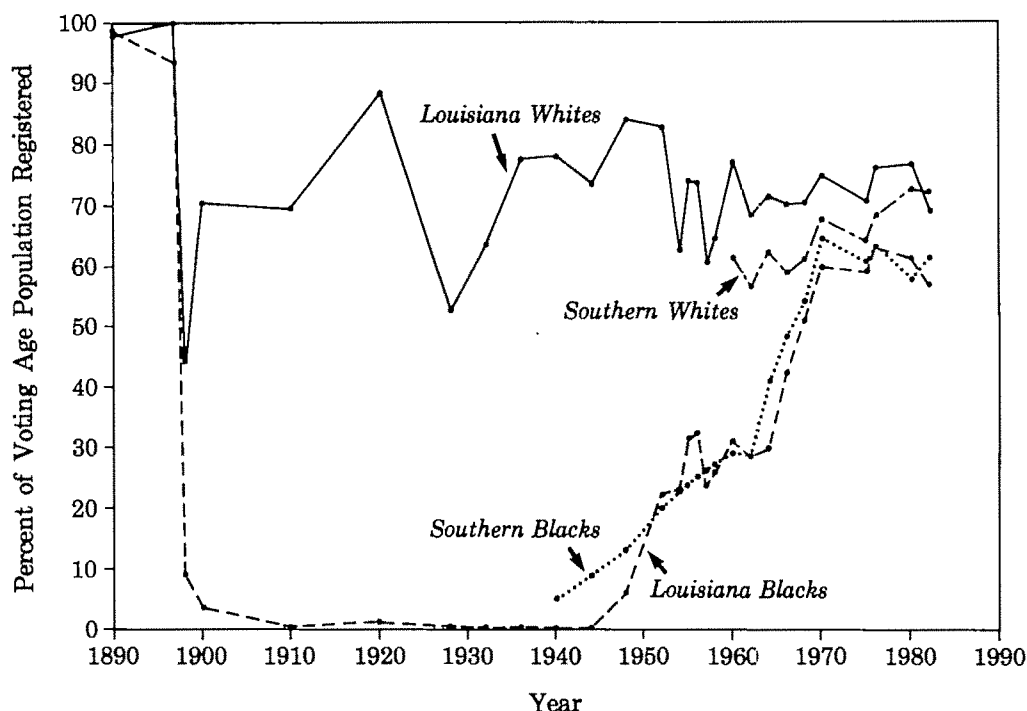
Fig. 1. Voter Registration Rates for Whites and Blacks in Louisiana and the South, 1890 to 1982
  Source: James (1981, p. 428)

Registration trends in Louisiana typified those in the other southern states (Myrdal 1944, pp. 486–90). New registration laws in 1898 and the disfranchising constitution of 1902 reduced the registration rate of Louisiana's black voters to less than 2 percent for 40 years. The white rate also suffered during this period, but by a smaller amount. For the South as a whole, the black registration rate averaged about 5 percent between 1900 and 1940, while the white rate probably ranged between 50 and 70 percent (Matthews and Prothro 1966, p. 17).

Beginning in the early 1940s, blacks had some success in transforming the racial state (Figure 1) in areas outside of the cotton-plantation regions. Registration drives in the larger southern cities produced significant gains after the Supreme Court outlawed the white primary in 1944, but stalled during the 1950s as white resistance stiffened in the deep South and in rural areas.

All of the essential features of the southern racial state were dismantled during the 1960s. Under massive pressure from the civil rights movement, the President and Congress strengthened the linkages between the national state and the local state. Vigorous federal enforcement of newly passed civil rights legislation and direct intervention by federal marshals and federalized national guardsmen reduced local-state autonomy.[7] The Voting Rights Act of 1965 (VRA) abolished local authority to exclude blacks from the franchise; literacy tests were outlawed and prior approval from the Depart-

---

[7] All of the mechanisms that reduce local-state autonomy were employed at one time or another during the civil rights struggles of the 1960s. Federal marshals and federalized national guardsmen were used on several occasions to desegregate southern public schools and universities. Local control over the appointment of officials was lost when a Federal Court forced a Florida sheriff to discharge a deputy who belonged to the KKK in St. Augustine (Friedman 1965, pp. 210–11). The most effective tool in forcing public-school desegregation was the withholding of federal funds provided to local schools under the Elementary and Secondary Education Act of 1965. National-state institutions such as the U.S. Commission on Civil Rights were created and others (such as the Office of Education and the Civil Rights Division of the Justice Department) were reorganized or expanded to enforce compliance with national policies (Orfield 1969).

ment of Justice or U.S. District Court for the District of Columbia was required before changing voting laws. Federal poll watchers were authorized, and the federal government was given the power to intervene directly to register blacks. Civil rights organizations led hundreds of voter-registration drives, first in the cities and then in the most rural areas of Mississippi. These efforts produced large increases in black rates (see Figure 1), which finally began to approach those of whites by the late 1960s (e.g., U.S. Commission on Civil Rights 1968; Lawson 1976). The following section analyzes the determinants of the racial state as it was being dismantled during the 1960s.

COTTON PLANTERS, WHITE FARMERS AND THE RACIAL STATE

Data and Measures

The difference between black and white enfranchisement rates indicates the state's racial bias. If no bias exists, blacks and whites should register at equal rates. If the state is biased against blacks, white rates should exceed black rates.

The percentage of the voting-age population registered to vote is a popular measure and simple to interpret, but percentages have two undesirable properties. They suffer from ceiling and floor effects when related to other variables in a causal analysis and do not express diminishing returns to scale. Logits correct both problems. The logit for blacks is

$$L(b) = \ln \{P(b) / [1 - P(b)]\} \quad (1)$$

where $P(b)$ is the registered proportion of the black voting-age population and ln indicates natural logarithms.

The degree of racial bias is the difference between the logits for blacks and whites,[8]

$$\Delta L = L(b) - L(w), \quad (2)$$

where $L(b)$ and $L(w)$ are computed for blacks and whites respectively using equation (1).

Equation (2) expresses a concept similar to the ratio of black to white voter-registration proportions, $P(b)/P(w)$. Both reflect the chances of a black person being registered compared to those of a white. A weighted least-squares routine was used to correct the regressions for heteroskedasticity.[9]

Appropriate voter-registration data were assembled on 575 of the 585 counties in the core southern states: Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, and South Carolina (James 1985). The ten missing counties had populations that were less than one percent black in 1960. Data were available for two panels before the VRA (1958 and 1964) and one after (1967) using information published by the U.S. Commission on Civil Rights (1959, 1961, 1963, 1968).[10] Most of the remaining data came from the U.S. Population and Agriculture Censuses or Matthews and Prothro (1963a, 1963b, 1966) and are discussed in James (1981, Appendix B).[11]

---

[9] The asymptotic variance of equation (1) is

$$V(L) = 1/NP(1 - P)$$

where N is the denominator of the proportion P (Theil 1970). Because the distribution of equation (1) is asymptotically normal, the variance of (2) is the sum of the variances given by $V(L)$ for blacks and whites minus twice the covariance. With independent processes, the covariance disappears asymptotically. Thus, the asymptotic variance of a linear combination of two logits is the sum of the variances. The regression weight for (2) is therefore the inverse of the sum of the individual variances given by $V(L)$ for blacks and whites.

[10] Counties are appropriate political units for testing the theory developed here. Of all of the different units of local government that might be analyzed, county-level governments have played the most important role in disciplining the southern rural black population. The U.S. Department of Agriculture farmer committee system is organized at the county level, as is the county sheriff's department, the most important law-enforcement agency in the rural South. While town and city governments gained authority over an increasing proportion of the black population as the rural-urban migration proceeded, relatively few southern counties have been dominated by large municipal governments. Even when municipal political structure is more important than county-level institutions, there is no reason to believe that it is constrained by class relations in a manner different from or contrary to that of the counties. If a county is dominated by the class structure of the cotton plantation, it is expected that the municipal and county governments would be similarly constrained. At any rate, it is not possible to distinguish municipal from nonmunicipal voter registration with the available data.

[11] Because census years did not correspond to the years for which registration data were available, values

---

[8] Readers of earlier versions of this paper suggested that black voter-registration rates are a better measure of the racial state than ratios or differences in black and white rates. I have resisted those suggestions because the available disfranchisement devices also disfranchised many whites. A contrast between the black and white rates is necessary to measure the degree of bias against blacks.

Black tenant farmers are the best indicator of the presence of labor-intensive plantation agriculture.[12] Plantations were owned by whites, but black tenants greatly outnumbered white farmers in areas where the plantation system was strongest. All white farm owners were actual or potential employers of black agricultural laborers, either as tenant farmers or wage workers. Thus, the presence of white farm owners, black farm tenants, and black farm laborers should be associated with a greater difference between white and black rates indicating greater racial bias in the local state. The other class categories should reduce the distance between white and black rates or have no effect.

The demographic weight of each class category is expressed as a proportion of the total county labor force. All class and occupational groups cannot be included in the regressions without creating perfect colinearity. Consequently, the effect of each class category must be interpreted in relation to the occupational groups excluded from the equation.

Because race-relations theories identify the white industrial working class as a primary carrier of racial discrimination, racial components of this class are included in the models. The textile and wood products industries employed more workers than any others in the South during the 1950s and 1960s. Black industrial workers were concentrated in the wood products industry, especially lumber, pulp wood, saw milling, and related forest-work that used unskilled manual labor released from farm work during the winter months. Whites in the wood industry were more likely to be furniture and fixtures

workers. The textile industry was an enclave for white workers from which most blacks were excluded before the 1960s (Northrup et al. 1970). Employees in other industries were lumped into a residual category, "other manufacturing workers," because of their smaller contribution to county employment levels.

The racial policies of large corporations are sometimes decisive for local race relations (e.g., Norgren 1967, pp. 465–69; Jacoway and Colburn 1982). The typical response of southern business leaders was "accommodation to what was perceived as inevitable change" (Jacoway 1982, pp. 8–14). Because racial data on the work force of large corporations were unavailable, I used the percentage of the total labor force employed in plants owned by firms on the *Fortune* list of the 500 largest corporations by sales. Local states were probably more sensitive to the political needs and preferences of national corporations as the number of jobs provided increased.

Organized labor is another national class actor that constrained local states. Labor unions often denied equal employment opportunities to blacks and opposed extending them civil rights at the community level (Hill 1977; Marshall 1965, 1967). Because county-level union membership data were unavailable, union strength was estimated by the count of AFL-CIO union locals present.

Variables tapping aspects of the social and economic structure of counties were included as suggested by Matthews and Prothro (1963a, 1963b, 1966) and others. These include median school years completed by race, median individual income by race, and whether a county had a black population majority. More education and higher incomes allegedly produced greater racial tolerance.

Race-relations theories predict that increasing black concentrations motivate whites to discriminate, which depresses black enfranchisement relative to white ones (e.g., Matthews and Prothro 1963a; Blalock 1967). On the other hand, resource-mobilization theories argue that population size facilitates the mobilization of insurgents (e.g., Tilly 1978; McAdam 1982; Daniel 1969). From this perspective, black majorities should be associated with greater black mobilization and smaller differences between black and white registration rates. Other measures of the mobilization of racial groups are dummy

---

for census variables were estimated for the panel years by linear interpolation. Other data were obtained from the U.S. Department of Labor, *Register of Reporting Labor Organizations* (1964) and Fortune *Plant and Product Directory* (1966) published by Time/Life Corporation. All variables correspond to the appropriate panel year unless otherwise indicated.

[12] A large number of regression models was estimated to identify the determinants of the farm-tenure variables used in the voter-registration equations reported above. Higher proportions of cotton and tobacco farms were very strongly related to tenant farmers of both races, strongly related to farm laborers of both races, and modestly related to the presence of white farm owners. Mechanization variables were negatively related to the presence of white and black tenant farmers but positively related to farm laborers of both races. Black farm owners were negatively related to mechanization but unrelated to cotton and tobacco farming.

variables indicating the presence (1 = present, 0 = absent) of organizations with explicitly racial purposes. Examples of black race organizations are the NAACP, CORE, SNCC, and the Urban League. The White Citizen's Councils and the Ku Klux Klan are typical white race organizations. These variables were collected by Matthews and Prothro (1966, pp. 164–65) for organizations operating in southern counties before February, 1961.

The Basic Class-Structure Model

Table 1 contains the results of regressing the difference between black and white voter-registration rate logits on the class-structure variables.[13] Black Farm Owners were associated with smaller differences between black and white rates (positive coefficients) in all three panels, although the 1964 coefficient is not significant. Black farmers had to deal with white merchants and bankers but were not dependent on the good will of white employers for their livelihood.

Other black agricultural workers were less autonomous than black farm owners and data in Table 1 reflect their vulnerability. Greater proportions of White Farm Owners, Black

---

[13] Negative coefficients imply a greater distance between black and white registration rates as the variable of interest increases. Positive coefficients imply diminishing distance between black and white rates.

Farm Tenants, and Black Farm Laborers increased the difference between white and black rates as indicated by negative coefficients (only the 1967 Black Farm Labor coefficient was not significant). These variables represented the presence of the classes with the greatest stake in black subordination. They had strong antidemocratic effects.

Competition theories argue that the white working class was the principal political enemy of southern blacks. Yet none of the remaining white class categories were linked to higher levels of discrimination. All of the white coefficients except those of White Farm Owners were positive or nonsignificantly different from the white middle-class variables excluded from the model. The white rural underclass and the white industrial working class were *not* the strongest supporters of the racial state.

The significantly large negative coefficients for Black Wood Workers can not be counted as supporting the competition thesis because blacks were confined to the most menial occupations in the wood products industry. Black wood workers were usually common laborers and wood cutters in the many small sawmills scattered across the South. The employment was seasonal, very labor intensive, and often merely an extension of agricultural employment.

Only three coefficients support the competition thesis: Black Other Manufacturing

Table 1. WLS Regressions of the Difference Between Black and White Registration Rate Logits on Class Variables: 575 Counties in Seven Southern States

	1958		1964		1967	
	b	s.e.	*b*	s.e.	*b*	s.e.
Labor Force Proportion that is						
Black Farm Owners	10.23*	3.60	2.90	5.61	20.85*	6.71
Black Farm Tenants	−15.08*	3.77	−17.25*	5.41	−44.83*	19.02
Black Farm Laborers	−6.36*	1.81	−7.96*	1.78	−1.74	1.59
Black Textile Workers	−10.09	10.54	−1.29	8.14	−10.64*	4.05
Black Wood Workers	−16.76*	3.06	−13.91*	4.04	−16.78*	3.99
Black Other Manuf. Workers	−10.46*	3.05	−6.14*	2.91	−2.08	2.25
White Farm Owners	−7.24*	1.63	−11.05*	2.00	−15.27*	2.08
White Farm Tenants	.91	4.77	.53	6.79	9.68	9.34
White Farm Laborers	7.32	5.02	20.90*	5.34	14.26*	4.62
White Textile Workers	−.73	.73	.05	.92	.43	.73
White Wood Workers	.76	2.32	−1.18	2.75	10.61*	3.26
White Other Manuf. Workers	3.62*	.03	1.42	1.02	.13	.83
Constant	−.99*	.13	−.89*	.13	−.41*	.11
R^2**	.42		.36		.30	

* Statistically significant at the .05 level, one-tailed test.
** R^2 values are for regressions on the weighted observations.

Table 2. WLS Regressions of the Difference Between Black and White Registration Rate Logits on Class and Social Structure Variables: 575 Counties in Seven Southern States

	1958		1964		1967	
	b	s.e.	b	s.e.	b	s.e.
Labor Force Proportion that is						
Black Farm Owners	12.11*	3.46	5.14	5.46	15.62*	6.55
Black Farm Tenants	−9.17*	3.69	−12.82*	5.27	−38.90*	17.99
Black Farm Laborers	−5.61*	1.89	−5.77*	2.09	−1.96	1.96
Black Textile Workers	−2.76	10.19	.24	7.87	−11.92*	3.88
Black Wood Workers	−16.73*	2.99	−12.34*	4.03	−14.84*	3.92
Black Other Manuf. Workers	−.95	3.20	.99	3.36	7.41*	2.57
White Farm Owners	−6.63*	1.78	−10.79*	2.20	−13.21*	2.21
White Farm Tenants	−.82	4.75	−2.37	6.90	9.78	9.40
White Farm Laborers	−2.94	4.91	9.72*	5.33	8.45*	4.51
White Textile Workers	−2.46*	.79	−.22	1.00	2.10*	.84
White Wood Workers	−.75	2.22	−3.96	2.70	6.96*	3.13
White Other Manuf. Workers	2.53*	1.10	3.35*	1.19	2.37*	.93
Fortune 500 Workers in 1965	.47	.51	−.44	.56	−1.52*	.42
AFL/CIO Locals in 1964 (10s)	−.06*	.01	−.01	.01	−.03*	.01
White Median School Years	−.14*	.05	−.06	.06	.22*	.06
Black Median Income ($1,000s)	.99*	.21	.72*	.19	.38*	.12
White Median Income ($1,000s)	−.05	.12	−.51*	.11	−.43*	.07
Black Majority Dummy	−.62*	.28	−.39	.31	.04	.30
Black Race Organization (1958)	.30*	.10	.29*	.10	.26*	.08
White Race Organization (1958)	−.46*	.09	−.35*	.08	−.13*	.07
Constant	−.19	.64	.04	.75	−2.52*	.78
R^2**	.50		.43		.40	

* Statistically significant at the .05 level, one-tailed test.
** R^2 values are for regressions on the weighted observations.

Workers in 1958 and 1964 and Black Textile Workers in 1967. They are discussed below.

Competition, Organization, and Status Effects

The models of Table 2 add measures suggested by other analysts to the basic model of Table 1.[14] None of the rural class coefficients were greatly affected. White Farm Owners, Black Tenant Farmers, Black Farm Laborers, and Black Wood Workers were still strongly related to racial bias in the extension of the franchise. Social-status and organization variables do not erase the effects of rural class structure.

Competition theories received less support. The Black Other Manufacturing Worker coefficients were insignificantly small in 1958 and 1964 and strongly positive in 1967 contrary to competition arguments. The White Textile Worker variable produced mixed results: strongly negative in 1958, which supports competition theory; insignificantly small in 1964 and strongly positive in 1967, which contradict competition theory. The negative Black Textile Worker coefficient in 1967 supports competition theory if the white working class is the discriminatory agent, but the positive 1967 White Textile Worker coefficient greatly reduces its impact.[15]

The 1967 *Fortune* 500 Workers coefficient is strongly negative, contradicting the claim

[14] Several possible models suggested by the literature were tested during the course of the research and rejected. Only the most parsimonious models are discussed above. The following variables were nonsignificant in all three panels and were omitted from the regression results presented to increase the precision of the estimates: Proportion Catholic of the Church Population in 1958, Black Median School Years, Log of Population Size. All three of these variables should have produced positive effects according to Matthews and Prothro (1963a).

[15] The effects of black and white textile employment essentially cancel each other when they are at their respective 1967 means of 0.015 and 0.082. The total effect of textile employment is $(-11.9 \times .015) = -.179$ for blacks plus $(2.1 \times .082) = 1.72$ for whites, equaling $-.007$, a negligible impact.

that national and multinational corporations were willing to accept black civil rights once racial change appeared inevitable (Jacoway and Colburn 1982; Cobb 1982). The anti-democratic effects of large corporations (Griffin, Wallace and Rubin 1986) appeared to increase rather than decrease after 1965 in these data.[16]

The effects of working-class organization were more consistent than the simple presence of workers. Larger numbers of AFL/CIO Union Locals had significantly negative effects in 1958 and 1967, but not in 1964. Southern unions, which were created and matured during the era of black disfranchisement and segregation, continued to threaten the job security and opportunities of black workers even after 1967 (Hill 1977). Black workers may have been reluctant voters where white workers were strongly organized into unions.

Black and White Race Organizations had significant effects in all three panels. Black organizations tended to equalize black and white rates, while white organizations produced an advantage for whites. Separate analyses not reported here suggest that larger population sizes and higher White Median School Years completed were the strongest positive determinants of White Race Organization in 1958.[17] Similarly, larger population sizes and higher Black Median School Years completed were the strongest positive determinants of Black Race Organization. White Farm Owners, Black Tenants, and Black Majorities, which occurred only in rural counties where the racial policies of white farmers and plantation operators were domi-

nant, decreased the likelihood of Black Race Organization. The working-class variables had very small or no effect on the mobilization of either race. Consequently, the Race Organization variables do not appear to be surrogates for white working-class mobilization against blacks.

As expected, communities with higher Black Median Incomes had more equal registration rates, but counties with high black incomes were rare. A few counties with cities such as Atlanta or Winston-Salem had black median incomes above $1,500 in 1958, but the average black median income for all 575 southern counties was $823. By contrast, white median income averaged $2,059. Black income is a function of the jobs available to blacks and, therefore, is highly related to county class structure. The lowest black median incomes were found in rural cotton and tobacco counties; the highest were in the cities offering industrial employment opportunities to blacks.

The white education and income status variables produced mixed results. White Median Income was strongly negative in 1964 and 1967. White Median School Years changed from negative in 1958 to positive in 1967. Two years of education were required to counteract the negative effects of $1,000 income in 1967 (0.43/0.22 = 1.95). The net results in the three panels suggest that *upper*-rather than lower-status groups were important carriers of racial discrimination, a result inconsistent with race-relations theories. Black Median School Years had no effect in any of the three panels and was omitted from the models reported.

---

[16] This negative coefficient may be an effect of workers rather than large corporations. The 1967 variables with the largest correlations with *Fortune* 500 Workers were White Textile Workers ($r = .26$) and White Other Manufacturing Workers ($r = .34$). In analyses not reported here, the coefficients for these two variables remained positive but dropped below statistical significance when *Fortune* 500 Workers were dropped from the model. The AFL/CIO union variable was only slightly related to *Fortune* 500 Workers ($r = .10$) and it had an independent negative effect. Thus, to the extent that the impact of *Fortune* 500 firms was due to workers in 1967, they had to be different than workers in smaller firms for reasons other than their membership in unions. The *Fortune* 500 Workers variable had no effect in 1958 and 1964; dropping it had no impact on the effects of other variables in the models for those years.

[17] Logit analyses of the determinants of both race organization variables were performed using the data set and are available from the author on request.

An Alternative Test of the Competition Thesis

Calculation of competition effects in the preceding section was based on the assumption that a class variable of either race could provide evidence favoring competition. Negative effects of white class variables were attributed to whites winning the competitive struggle with blacks. Negative effects of black class variables were attributed to blacks losing the competition with whites.

Popular theories argue that competition stems from racial *compositions,* not levels of relevant populations (Blalock 1967; Wilson

Table 3. WLS Regressions of the Difference Between Black and White Registration Rate Logits on Class, Social Structure, and Interaction Variables: 575 Counties in Seven Southern States

	1958		1964		1967	
	b	s.e.	b	s.e.	b	s.e
Labor Force Proportion that is						
Black Farm Owners	6.94*	4.11	−9.99	6.71	9.68	7.30
Black Farm Tenants	−8.77*	3.68	−3.98	5.92	−45.17*	18.17
Black Farm Laborers	−5.64*	1.88	−3.64	2.60	−1.11	1.99
Black Textile Workers	−1.32	10.17	5.93	7.90	−10.44*	3.95
Black Wood Workers	−16.78*	2.98	−10.42*	3.99	−14.44*	3.95
Black Other Manuf. Workers	−1.85	3.21	.69	3.34	12.45*	4.03
White Farm Owners	−5.11*	1.90	−7.48*	2.33	−11.43*	2.32
White Farm Tenants	−.71	4.73	−10.10	7.17	9.49	9.39
White Farm Laborers	−2.65	4.89	7.40	5.58	5.96	4.59
White Textile Workers	−2.41*	.78	−.44	.99	1.67*	.86
White Wood Workers	−1.21	2.23	−4.89*	2.68	6.05*	3.13
White Other Manuf. Workers	2.92*	1.11	3.23*	1.19	1.12	1.23
Fortune 500 Workers (1965)	.47	.51	−.57	.55	−1.45*	.42
AFL/CIO Locals in 1964 (10s)	−.06*	.01	−.01	.01	−.02*	.01
White Median School Years	−.16*	.05	−.06	.06	.19*	.06
Black Median Income ($1,000s)	1.18*	.23	.74*	.20	.43*	.12
White Median Income ($1,000s)	−.04	.12	−.41*	.11	−.43*	.07
Black Majority Dummy	−.62*	.28	−.25	.31	.20	.30
Black Race Organization (1958)	.31*	.10	.28*	.10	.27*	.08
White Race Organization (1958)	−.50*	.09	−.35*	.09	−.16*	.07
Ratio of Black to White						
Farm Owners	.53*	.23	1.09*	.26	.52*	.27
Farm Tenants			−.12*	.04		
Farm Laborers			−.06*	.03		
Other Manuf. Workers					−.69*	.37
Constant	−.37	.64	−.25	.74	−2.15*	.79
R^{2}**	.50		.45		.40	

* Statistically significant at the .05 level, one–tailed test
** R^2 values are for regressions on the weighted observations.

1978).[18] A composition measure was obtained by dividing the black proportion of the labor force by the white proportion. For example, a measure of the competition between black and white textile workers is the county ratio of black to white textile workers. The evidence for the following analyses will

---

[18] All class variables were calculated as proportions of the total labor force, but an *increase* in the proportion of white textile workers, for example, does not imply that the black proportion will *decrease* in a complementary fashion. Because significant portions of the labor force are in the excluded occupations, the white and black portions of a particular industry can rise (or fall) simultaneously. These variables, scaled to the size of the county labor force, produce level effects. Composition arguments focus on the effect of racial components of particular segments of the population relative to each other. For example, the effect of the number of black textile workers relative to the number of white textile workers is examined.

be judged to favor competition theories if the total effect of a black class variable or an interaction is negative. This approach biases the analysis in favor of competition theories (See Appendix).

The six possible ratio interactions between black and white class variables were added to the basic equations of Table 2 one at a time. Only those models that produced statistically significant results were presented in Table 3. No interaction effects were significant for Textile Workers or Wood Workers in any of the three panels. The Farm Owners interactions were significant but positive in all three panels. Furthermore, the total effects of Black Farm Owners in 1958 and 1967 were positive for all levels of White Farm Owners. Black Farm Owners produced negative effects in 1964 but only if the concentration of White Farm Owners was greater than 10.9 percent.

Whites were supposedly more vulnerable to competition when their concentrations were small, yet white farmers averaged 8.7 percent of the county labor force and blacks only 1.7 percent. Thus, little or no support for competition theories was produced by the Farm Owners, Textile Workers, or Wood Workers variables using either criterion described in the Appendix.

The Farm Tenants and Farm Laborers interactions were both negative in 1964, as were the total effects of Black Farm Tenants and Farm Laborers. The Other Manufacturing Workers interaction was negative in 1967 and the total effect of Black O. M. Workers was negative when White O. M. Workers were less than 5.5 percent. These results were consistent with competition theories, but the overall support for competition was weak. Evidence for class domination of the local state was much stronger than the evidence for competition theories, both within and across panels, regardless of the measure of competition employed.

CONCLUSION

Theories of the state and race-relations theories are usually discussed as if they were unrelated to each other. Theories of the state focus on the national state, ignore local political institutions, and fail to account for the variation in the local implementation of national policies. Race-relations theories ignore the effects of national- and local-state structures altogether, suggesting instead that status distinctions among whites or labor-market competition between blacks and whites are the primary determinants of racial discrimination. This research suggests that both sets of theories need reexamination.

First, local-state structures have a degree of autonomy from the national state, especially in federal systems like the United States. The central state's capacity to enforce compliance with national policies is reduced to the extent that local states have the power to appoint local officials and to raise revenues. Furthermore, local states are strongly constrained by local class structures. Insurgent movements based in local class structures may find the local state to be more vulnerable to democratic pressure than the national state under some circumstances (e.g., Dearlove and Saunders 1984; Rhodes 1981), but local states also provide opportunities for oligarchs.

White planters and farmers in the American South were able to achieve regional hegemony through their domination of the local state when they were no longer able to contend for national power.

Second, state structures are not neutral mediators of pluralist combat among competing interest groups. Local-state structures in the American South were neither liberal nor democratic before the Civil War, nor were they democratic before the civil rights victories of the 1960s. Instead, the local state in the American South was a racial state that imposed political and social burdens on blacks and conferred advantages on whites.

The southern racial state was not an accident of history but was created by white plantation owners, white farm owners, and their allies to discriminate against blacks. The fortunes of white farmers and planters were strongly dependent on the political impotence and docility of black agricultural workers. Disfranchisement removed blacks as a political force. State-enforced segregation imposed social and economic hardships. Violence and coercion, often performed by state agents and always accomplished with state consent, reduced the aspirations of black workers and increased their vulnerability to the authority of white employers. Landlords and employers were granted special privileges denied to employees. The racial state ensured that black workers were the most disadvantaged of all employees.

Cotton planters were the strongest defenders of the antebellum slave state. Cotton planters and white farmers were the chief architects and beneficiaries of the racial state during the late nineteenth century. When the racial state was under massive attack by the civil rights movement of the 1960s, it remained strongest in those areas where white farm owners and planters dominated local class structures. No other variables were as strongly or consistently related to the difference between black and white voter-registration rates as were the combined effects of the concentration of white farm owners and the black farm tenants and laborers who typically worked for white farm owners. The effects attributable to status distinctions, labor-market competition, and organizational capacities did not eliminate the strong direct effects of class structure.

Class and political structures establish the social context within which organizations and

status distinctions produce their effects. Black race organizations had a democratizing effect; white race organizations were antidemocratic. Labor unions also appear to have had antidemocratic effects. The same local political structure that stripped blacks of basic citizenship rights also denied them equal education and restricted their access to manufacturing jobs. That racial conflict occurred among status groups or between segments of the working class from time to time is not surprising given this context. *To elevate those conflicts to the status of primary determinant of southern race relations is to confuse the proximate cause of conflicts with the fundamental cause.*

Theories of the state and theories of American race relations could be improved by consideration of the effects of local-state structures. Existing theories of the state fail to account for the great variation in the local enforcement of national civil rights policies during the 1960s. Race-relations theories ignore the effects of local-state structures that enforced racial discrimination and maintained southern racial politics. Local-state theory addresses both problems.

APPENDIX

Race-relations theories argue that racial competition is a function of the racial composition of populations; higher proportions of blacks relative to whites produce higher levels of discrimination against blacks. A simple model is

$$Y = b_0 + b_1X_1 + b_2X_2 \\ + b_3X_3 \\ + b_4X_1/X_2 + e \qquad (3)$$

where Y is a measure of racial bias, X_1 and X_2 are black and white components of the labor force, X_3 some other independent variable of interest, and e is an error term. If b_4 equals zero, the effects of X_1 and X_2 are additive; there is no effect of composition. Competition is absent.

When b_4 is not equal to zero, the effect on Y of changes in the black labor-force component X_1 is obtained by taking the partial derivative of equation (3) with respect to X_1 (Stolzenberg 1979) which yields

$$\partial Y/\partial X_1 = b_1 + b_4/X_2 \qquad (4)$$

The effect of changes in X_1 on Y has two components: one is a function of changes in the level of X_1, and the other is a function of changes in the level of X_1 relative to the level of X_2.

Different interpretations of equation (4) are possible. The first treats b_4 as a measure of competition because it indicates the effect of X_1 relative to X_2.[19] Negative values

of b_4 indicate that whites have won the competitive struggle with blacks. The second finds support for competition if the total effect of X_1 is negative and the ratio interaction is non-zero. In other words, if the total effect of an increase in X_1 is negative and non-linear because of the relative sizes of X_1 and X_2, it is attributed to the competitive advantage of whites over blacks.[20]

REFERENCES

Alford, Robert R. and Roger Friedland. 1985. *Powers of Theory*. Cambridge: Cambridge University.

Alston, Lee J., and Joseph P. Ferrie. 1985. "Labor Costs, Paternalism, and Loyalty in Southern Agricultural." *Journal of Economic History* 45:96–117.

Badger, Anthony J. 1980. *Prosperity Road: The New Deal, Tobacco, and North Carolina*. Chapel Hill: University of North Carolina.

Barkan, Steven E. 1984. "Legal Control of the Southern Civil Rights Movement." *American Sociological Review* 49:552–65.

Becker, Gary. 1971. *The Economics of Discrimination*, 2nd ed. Chicago: University of Chicago.

Bendix, Reinhard. 1964. *Nation-Building and Citizenship*. Berkeley: University of California.

Bensel, Richard Franklin. 1984. *Sectionalism and American Political Development 1880–1980*. Madison: University of Wisconsin.

Blalock, Hubert M., Jr., 1967. *Toward a Theory of Minority Group Relations*. New York: Wiley.

Block, Fred. 1977. "The Ruling Class Does Not Rule: Notes on the Marxist Theory of the State." *Socialist Revolution* 33:6–28.

Bloom, Jack M. 1987. *Class, Race, and the Civil Rights Movement*. Bloomington: Indiana University.

Cobb, James C. 1981. *The Selling of the South*. Baton Rouge: Louisiana State University.

Cohen, William. 1976. "Negro Involuntary Servitude in the South, 1865–1940." *Journal of Southern History* 42:31–60.

Daniel, Johnnie. 1969. "Negro Political Behavior and Community Political and Socioeconomic Structural Factors." *Social Forces* 47:274–80.

Daniel, Pete. 1972. *The Shadow of Slavery: Peonage in the South 1901–1969*. Urbana: University of Illinois.

Davis, Allison, Burleigh B. Gardner, and Mary R. Gardner. 1941. *Deep South*. Chicago: University of Chicago.

Dearlove, John, and Peter Saunders. 1984. *Introduction to British Politics*. Cambridge: Polity Press.

Fligstein, Neil. 1981. *Going North*. New York: Academic Press.

---

[19] This interpretation is analogous to the case noted by Stolzenberg (1979, p. 466), in which the second-order

term in a parabolic regression is viewed as a measure of convexity or concavity.

[20] An anonymous reader of an earlier version of this essay suggested a third interpretation of equation (4). In this view, the total effect on Y is interpreted in terms of levels of X_1 and X_2 because, it is argued, no causal autonomy can be attributed to the two components of (4). This position appears unwarranted when theories are strong enough to allow specification of the functional form of the determination of Y by the Xs, as is the case with race-relations theories. The failure to find empirical support for that form should count against the theory, but see Stolzenberg (1979, p. 472) for an opposing position.

Friedman, Leon. 1965. "The Federal Courts of the South." Pp. 186–213 in *Southern Justice*, edited by Leon Friedman. New York: Pantheon.

Griffin, Larry J., Michael E. Wallace, and Beth Rubin. 1986. "Capitalist Resistance to the Organization of Labor Before the New Deal: Why? How? Success?" *American Sociological Review* 51:147–67.

Hahn, Steven. 1983. *The Roots of Southern Populism*. New York: Oxford.

Hannan, Michael T. 1979. "The Dynamics of Ethnic Boundaries in Modern States." Pp. 253–75 in *National Development and the World System*, edited by John W. Meyer and Michael T. Hannan. Chicago: University of Chicago.

Harlan, John Marshall. [1896] 1970. "*Plessy v. Ferguson,* (Justice Harlan, dissenting)." Pp. 150–52 in *The Civil Rights Record*, edited by Richard Bardolph. New York: Thomas Y. Crowell.

Hill, Herbert. 1977. *Black Labor and the American Legal System*. Washington, DC: Bureau of National Affairs.

Hine, Darlene Clark. 1979. *Black Victory: The Rise and Fall of the White Primary in Texas*. Millwood, NY: KTO Press.

Hoffman, Erwin D. 1969. "The Genesis of the Modern Movement for Equal Rights in South Carolina, 1930–1939." Pp. 193–214 in *The Negro in Depression and War*, edited by Bernard Sternsher. Chicago: Quadrangle Books.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. Berkeley: University of California.

Jacoway, Elizabeth. 1982. "Civil Rights and the Changing South." Pp. 1–15 in *Southern Businessmen and Desegregation*, edited by E. Jacoway and D.R. Colburn. Baton Rouge: Louisiana State University.

Jacoway, Elizabeth, and David R. Colburn, eds. 1982. *Southern Businessmen and Desegregation*. Baton Rouge: Louisiana State University.

James, David R. 1981. "The Transformation of Local State and Class Structures and Resistance to the Civil Rights Movement in the South." Ph.D. diss., University of Wisconsin, Madison.

———. 1985. *Southern Voter Registration Study* [MRDF]. Bloomington: Institute for Social Research, Indiana University.

———. 1986. "Local State Structure and the Transformation of Southern Agriculture." Pp. 150–78 in *Studies in the Transformation of United States Agriculture*, edited by Eugene Havens with Gregory Hooks, Patrick Mooney, and Max Pfeffer. Boulder, CO: Westview Press.

Key, V.O. 1949. *Southern Politics in State and Nation*. New York: Vintage.

Kirby, Jack Temple. 1987. *Rural Worlds Lost: The American South 1920–1960*. Baton Rouge: Louisiana State University.

Kousser, J. Morgan. 1974. *The Shaping of Southern Politics*. New Haven, CT: Yale.

Lawson, Steven F. 1976. *Black Ballots: Voting Rights in the South, 1944–1969*. New York: Columbia University.

LeRay, Jr., Nelson and Grady B. Crowe. 1959. "Labor and Technology on Selected Cotton Plantations in the Delta Area of Mississippi, 1953–1957." Mississippi Agricultural Experiment Station Bulletin 575.

Lipsky, Michael. 1980. *Street-Level Bureaucracy*. New York: Russell Sage.

Litwak, Leon. 1980. *Been in the Storm So Long*. New York: Vintage.

Mandle, Jay R. 1978. *The Roots of Black Poverty*. Durham, NC: Duke University.

Marshall, F. Ray. 1965. *The Negro and Organized Labor*. New York: Wiley.

———. 1967. *Labor in the South*. Cambridge: Harvard University.

Marx, Karl. [1843] 1978. "On the Jewish Question." Pp. 26–52 in *The Marx-Engels Reader*, 2nd ed., edited by Robert C. Tucker. New York: W.W. Norton.

Matthews, D.R. and J.W. Prothro. 1963a "Social and Economic Factors and Negro Registration in the South." *American Political Science Review* 57: 24–44.

———. 1963b. "Political Factors and Negro Voter Registration in the South." *American Political Science Review* 57: 355–367.

———. 1966. *Negroes and the New Southern Politics*. New York: Harcourt, Brace, and World.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency 1930–1970*. Chicago: University of Chicago.

Meier, August and Elliot Rudwick. 1976. *From Plantation to Ghetto*, 3rd ed. New York: Hill and Wang.

Myrdal, Gunnar. 1944. *An American Dilemma*. New York: Harper and Row.

Nagel, Joane, and Susan Olzak. 1982. "Ethnic Mobilization in New and Old States: An Extension of the Competition Model." *Social Problems* 30: 127–43.

Nielsen, Francois. 1985. "Toward a Theory of Ethnic Solidarity in Modern Societies." *American Sociological Review* 50:133–49.

Norgren, Paul H. 1967. "Fair Employment Practice Laws—Experience, Effects, Prospects." Pp. 541–70 in *Employment, Race, and Poverty*, edited by A.M. Ross and H. Hill. New York: Harcourt, Brace, and World.

Northrup, Herbert R., R.L. Rowan, D.T. Barnum, and J.C. Howard. 1970. *Negro Employment in Southern Industry*. Philadelphia: University of Pennsylvania.

Orfield, Gary. 1969. *The Reconstruction of Southern Education*. New York: Wiley-Interscience.

———. 1978. *Must We Bus: Segregated Schools and National Policy*. Washington, DC: The Brookings Institution.

Paige, Jeffery M. 1975. *Agrarian Revolution: Social Movements and Export Agriculture in the Underdeveloped World*. New York: Free Press.

Pederson, Harald A. and Arthur F. Raper. 1954. "The Cotton Plantation in Transition." *Mississippi Agricultural Experiment Station Bulletin* 508.

Peterson, Paul E. 1981. *City Limits*. Chicago: University of Chicago.

Piven, Frances Fox and Richard A. Cloward. 1971. *Regulating the Poor*. New York: Vintage.

Przeworski, Adam. 1985. *Capitalism and Social Democracy*. Cambridge: Cambridge University.

Rabinowitz, Howard. 1978. *Race Relations in the Urban South, 1865–1890*. New York: Oxford.

Ransom, Roger L. and Richard Sutch. 1977. *One Kind of Freedom*. London: Cambridge University.

Rhodes, R.A.W. 1981. *Control and Power in Central-Local Government Relations*. Hants, Great Britain: Gower Publishing Company.

Salamon, Lester M. and Stephen Van Evera. 1973. "Fear, Apathy, and Discrimination: A Test of Three Explanations for Political Participation." *American Political Science Review* 67:1288–1306.

Schwartz, Michael. 1976. *Radical Protest and Social Structure*. New York: Academic Press.

Shifflett, Crandall A. 1982. *Patronage and Poverty in the Tobacco South*. Knoxville: University of Kentucky.

Skocpol, Theda. 1979. *States and Social Revolutions*. Cambridge: Cambridge University.

St. Angelo, Douglas and Paul Puryear. 1982. "Fear, Apathy, and Other Dimensions of Black Voting." Pp. 109–30 in *The New Black Politics*, edited by Michael B. Preston, L.J. Henderson, Jr., and P. Puryear. New York: Longman.

Stampp, Kenneth M. 1956. *The Peculiar Institution*. New York: Vintage.

Stolzenberg, Ross M. 1979. "The Measurement and Decomposition of Causal Effects in Nonlinear and Nonadditive Models." Pp. 459–88 in *Sociological Methodology 1980*, edited by Karl F. Schuessler. San Francisco: Jossey-Bass.

Theil, Henri. 1970. "On the Estimation of Relationships Involving Qualitative Variables." *American Journal of Sociology* 76:103–54.

Tilly, Charles. 1978. *From Mobilization to Revolution*. Reading, MA: Addison-Wesley.

U.S. Commission on Civil Rights. 1959. *Report*. Washington, DC: U.S. Government Printing Office.

———. 1961. *Voting: U.S. Commission on Civil Rights Report, Book 1*. Washington, DC: U.S. Government Printing Office.

———. 1963. *Civil Rights '63*. Washington, DC: U.S. Government Printing Office.

———. 1965. *Voting. Volume 1 of Hearings Held in Jackson, Mississippi. February 16–20, 1965*. Washington, DC: U.S. Government Printing Office.

———. 1968. *Political Participation*. Washington, DC: U.S. Government Printing Office.

Wiener, Jonathan M. 1978. *Social Origins of the New South*. Baton Rouge: Louisiana State University.

Williams, Walter. 1980. *Government by Agency: Lessons from the Social Program Grants-in-Aid Experience*. New York: Academic Press.

Wilson, William J. 1978. *The Declining Significance of Race*. Chicago: University of Chicago.

Woodman, Harald D. 1977. "Sequel to Slavery: The New History Views the Postbellum South." *Journal of Southern History* 43:523–54.

Woodward, C. Vann. 1974. *The Strange Career of Jim Crow*, 3rd ed. New York: Oxford.

Woofter, Jr., T.J. 1936. *Landlord and Tenant on the Cotton Plantation*. WPA Monograph V. Washington, DC: U.S. Government Printing Office.

Wright, Erik Olin. 1978. *Class, Crises, and the State*. London: New Left Books.

Wright, Gavin. 1986. *Old South, New South*. New York: Basic Books.

MANUSCRIPTS FOR THE
ASA ROSE SOCIOLOGY SERIES

Manuscripts (100 to 300 typed pages) are solicited for publication in the *ASA Arnold and Caroline Rose Monograph Series*. The Series welcomes a variety of types of sociological work — qualitative or quantitative empirical studies, and theoretical or methodological treatises. An author should submit three copies of a manuscript for consideration to the Series Editor, Professor Teresa A. Sullivan, University of Texas, Population Research Center, 1800 Main, Austin, TX 78712.

SECRECY AND STATUS: THE SOCIAL CONSTRUCTION OF FORBIDDEN RELATIONSHIPS*

LAUREL RICHARDSON
The Ohio State University

This paper discusses how status and secrecy affect the social construction of secret, forbidden, intimate relationships. Based on an analysis of 65 intensive interviews with single women involved with married men, two stages of intimacy emerge: (1) exchanging secrets about the self or "Becoming Confidantes" and (2) creating mutual secrets or "Becoming a We." The man's marital status, reinforced by his gender and socioeconomic status, has major effects concerning time constraints, expectations of temporariness, and privacy. These lead to intense feelings, idealization, and trust, which enhance the woman's commitment to perpetuating the relationship. The relationship is perpetuated through the construction of mutual secrets (rituals and property), which are imbued with intense symbolic significance. The strategies used to conceal the relationship increase the woman's dependence on it and reduce her power within it. Secrecy protects the interests of the powerful.

INTRODUCTION

Eighty years ago, Georg Simmel proposed that all social relationships "can be characterized by the amount and kind of secrecy within them and around them" (1950, p. 331). Yet, how status and secrecy affect the construction of social relationships has been little studied. The purpose of this paper is to theorize how status and secrecy affect the construction of a particular category of relationships: secret, forbidden, sexual relationships.

The secret, to paraphrase Simmel (1950, p. 330), was one of the major achievements of humankind because it permitted an immense enlargement of the world, the possibility of hiding reality and creating a second world alongside the manifest one. The secret, he argued, is a general sociological form of major significance, regardless of its content. Because some measure of secrecy exists in all relationships, ignoring it limits our understanding of how social relationships are constructed and maintained.

Analytically, Simmel distinguished two kinds of secrets. There are secrets about the self that are privately constructed and held and secrets about a social unit that are socially constructed and held, such as the private language of twins, the plots of co-conspirators, or the very existence of the co-conspirators. Secrets are possessions whose value is enhanced by denying knowledge of them to others. If secrets are shared, there is the risk of betrayal; if they are not shared, relationships are circumscribed. Following Simmel, understanding how secrecy is managed can tell us a great deal about how relationships are constructed. However, Simmel did not focus on the issue that concerns us here, namely, how status imbalances affect the continuing social construction of secret relationships.

A major proposition of the social constructionist school is that the meanings attached to relationships vary according to status and that status differentials have consequences for the relationship's construction. Interactionally, status differentials are *power* differentials, and face-to-face interactions reflect the constraints or freedoms of relative social positions (cf. Karp and Yoels 1979, pp. 124–81). Constructionists note that we simultaneously inhabit a micro world of face-to-face relationships and a macro world of abstract and remote institutional arrangements. The micro world makes full sense only in the context of the macro one that enfolds it (cf. Berger and Berger 1975, p. 8). Analyses of face-to-face relationships must take account of broader institutional arrangements, including those of status differentials (cf. Reynolds 1987).

Secret *intimate* relationships are particularly salient for an analysis of secrecy and

status in relationship construction. In secret relationships, ample opportunity exists both to reveal secrets about the self and to conceal one's biography, identity, and purposes. Second, ample opportunity exists to create mutual secrets, including the very existence of the relationship. Moreover, because intimate relationships are privatized, secret ones are doubly removed from the routine world. In such situations, restraints are minimized, so the impact of status is in greater relief.

One category of secret relationships, forbidden sexual relationships, is arguably the largest and most significant class of secret relationships. More individuals are likely to be so involved and more individuals are likely to be deeply affected by their own or significant others' involvements than any other kind of secret relationship. For example, somewhere between 18 percent and 32 percent of single women become involved with married men (Richardson 1986); approximately 50 percent of husbands and 35 percent of wives have extramarital relationships (Atwater 1982; Reiss, Anderson, and Sponaugle 1980); approximately 25 percent of adult women were sexually molested by adult men when they were girls (under 16) (cf. Russell 1986). Untold numbers of professors become involved with students, doctors with patients, ministers with parishioners, employers with employees. The emotional significance of these relationships can be intense and long-lasting. Women sexually harassed at work experience extreme stress (MacKinnon 1979). Single women involved with married men often find themselves more emotionally committed than they had expected to be (Richardson 1985). And childhood sexual experiences, including incest, have major, long-lasting emotional consequences (Koch 1980).

A feature of cross-sex secret, forbidden sexual relationships is that they are typically constructed between *status unequals*. One status imbalance usually exists: age, class, or marital status. Consequently, the construction of cross-sex secret liaisons should reflect the major consequences of status differences, *power* differences. Secret sexual liaisons are likely to be *power-imbalanced* relationships.

Liaisons between single women and married men should provide the paradigmatic case for these secret, forbidden, power-imbalanced relationships, not only because of their numerical abundance, but for cultural-

historical, interactional, and structural reasons. Adulterous relationships provide a world in which self-disclosures and mutually created secrets can proliferate. De Rougemont (1956), in his review of love in the western world, argued that there is an inescapable conflict between passion and marriage in the West, with adultery nearly synonymous with passion. He speculated that westerners are drawn to adultery, an act that destroys marital happiness, as much as they are to acts which ensure marital happiness, because marriage and ardor are in opposition.

Because of social-structural and social-psychological reasons, men have greater power than women. Men are more highly valued, have better jobs, and earn more than women (cf. Reskin and Hartmann 1986). The male and the masculine are viewed as superior to the female and the feminine, and a double standard exists such that differential rewards and sanctions are given to men and women for the same behavior (cf. Eichler 1980). Further, gender differences exist regarding the meaning of love and the importance of intimate relationships (cf. Gilligan 1982; Hochschild 1983; Diamond 1983). Because women tend to value relationships more highly than men, work more to sustain them, and feel more responsible for their outcomes, women are likely to care more about having and preserving a relationship than men are. The "principle of least interest" holds (Waller 1938; Blau 1964). Whoever has the lesser interest in maintaining a relationship has the greater power within it.

In single woman/married man liaisons, the male has greater structural and social-psychological resources because of his marital status, as well. Being married confers greater prestige than being single, gives a veneer of stability and access to a network of couples (Stein 1981), as well as an immediate alternative sexual partner. The effect of marital status difference on the construction of the relationship can be set in relief. Since the men in these liaisons are usually of a higher socioeconomic class than the women, there are also socioeconomic status differences. But, in addition, because socioeconomic peers are getting involved (Richardson 1979; 1985), how socioeconomic inequalities interact with differences in gender and marital status can be explored.

Based on intensive interviews with single women involved with married men, I gener-

ate theories about the links between secrecy and status in the social construction of forbidden sexual liaisons. Two sequential stages of intimacy construction emerge and demand analysis: (1) exchanging secrets about the self, or *Becoming Confidantes;* and (2) creating mutual secrets, or *Becoming a We.* My analysis focuses on how the clandestine nature of these liaisons transforms, heightens, and/or intensifies normative social expectations and structures, including the impact of status differences. I find that the man's marital status, reinforced by his gender status and superior socioeconomic status, has major and significant consequences for the construction of the liaison.

PROCEDURES

The data come from intensive interviews with 65 single women who had or were having long-term (over a year) intimate relationships with married men. The interviewees came from three sources. First, whenever I attended a public or social occasion, I mentioned the project, and women I met in these settings volunteered to be interviewed. Second, a respondent would occasionally suggest other potential interviewees. Third, women, hearing of the project, requested to be interviewed.

The frequently reinstituted "snowball" sampling procedure generated an extensive list of potential interviewees from a wide variety of social networks. After 48 interviews with a diverse set of women over a period of five years, saturation (cf. Glaser and Strauss 1967) was experienced. Since I was fairly accurately predicting the women's accounts, I discontinued interviewing. During the analysis and writing stages, though, I followed 17 women who contacted me during the early stages of their liaisons. I interviewed them at approximately three-month intervals over a minimum period of 12 months. Because these women's accounts were substantially in agreement with those of other women I interviewed, the usual issues regarding the reliability and validity of retrospective biography are somewhat muted.

The women resided in rural and urban areas in all regions of the country. When their relationships began, they ranged in age from eighteen to fifty-six with a median age of twenty-eight. Sixteen women were high school graduates, thirty-six were college

graduates, and fourteen had advanced degrees. Nine were students at some point during their liaisons. Occupationally, they held a variety of blue-collar, pink-collar, clerical, administrative, and professional positions. Twenty-seven of them had never been married, thirty-eight were divorced. Fifteen were raising children. Fifty-three were younger and less established than the men; eighteen had been supervised or mentored by their lovers at some point during their relationships; seven of them were occupational peers; and five were socioeconomically superior to their lovers. When the relationships began, all of the husbands were living with their wives, and fifty-two of them had minor or college-aged children. None of the husbands claimed to be in "open marriages," although five explicitly said their wives overlooked their infidelity. The relationships studied lasted from one to twenty-five years, and, more than half were over at the time of, or shortly after, the interview.

The interviews were conducted in private settings such as the respondent's home or mine, a private office, or a hotel room. An open-ended format elicited detailed information on the woman's feelings, attitudes, and behaviors during different phases of the liaison. Interviews were tape recorded and transcribed.

Because the sample of women was not randomly drawn, precise representativeness is unknown. The sample, however, is heavily weighted toward women with higher education and white-collar occupations. The analysis searches for patterns and themes in the women's accounts, and presents these as processes and types. Because the women's accounts make sense within the growing substantive and theoretical understanding of women's experiences, considerable face validity can be claimed for the study. Because the relations are long-term ones, however, the analysis is not generalized to one-night stands or abbreviated affairs.

I looked at how features of these liaisons resemble features of normatively approved relationships and how the forbiddenness and secrecy of the liaisons transform or intensify those features. My conceptualizations are grounded in the data, and theoretical ideas are inductively generated (Glaser and Strauss 1967). All generalizations are based on the dominant pattern or the clear majority of the respondents. When I quote from a particular

interview, it represents a common interview theme. Since the interviews were fluid and open ended, not all respondents volunteered all themes. However, I do not discuss themes of processes that are not general.

STAGE ONE: BECOMING CONFIDANTES

Building an intimate relationship is an accomplishment that is affected by the context in which the intimacy is being built. Yet, little attention has been paid to the structural contexts that impede or enhance self-disclosures and intimacy (but, notably, see Davis 1973). I find that the man's marital status creates a context of privacy, time constraints, and expectations of temporariness that encourage revealing secrets about the self. Revealing them increases the woman's trust in and idealization of her lover and their relationship, which, in turn, intensifies the woman's commitment. The intimacy process is a dynamic one greatly affected by the fact the involved man is a *married* man.

Single women see the beginning and middle phases of their relationships with married men as "total," "emotionally pure," and "uniquely intimate." All the women I interviewed described their lovers as special people. They were seen as "compassionate and exciting," "like no man ever known before," and "every woman's dream." Even women who did not expect to adulate their lovers, women who were "career driven" or usually "emotionally reserved" commented on the "truly special qualities" of their lovers, their "generosity," "openness," "strength of character," and "spontaneity."

The pattern of idealization suggests that sociological factors of the relationships are operative, rather than the psychological characteristics of the women. The relationships have in common that they begin as a secret and that the man is married, the woman is not. Because the man is married, the relationship's existence is kept secret from particular significant others, such as her children, her employer, and his wife (Richardson 1985). Because of his marital status, moreover, three structural conditions arise: privacy, time constraints, and expectations of nonpermanence. It is precisely these conditions that are structurally conducive to the woman's idealization of her lover and the relationship.

First, because the man is married, activities that sustain the relationship are carried on predominantly in *privacy*. The world-out-there, the normative social structure with its roles and rules, expectations and obligations, can be laid aside as the couple constructs a world-in-here, a world freer of normative social constraint and cultural definitions. Alone-together, the pair can construct the relationship, its boundaries, and its focus. Between them, they have the resources, such as their incomes and her home, to create and sustain the requisite privacy. Privacy leads subsequently to greater relational autonomy. As one woman summarized it, "We didn't have expectations of each other, and we didn't have to exploit each other. The relationship didn't have to have a goal. We could just be."

Second, because he is married, the relationship's *time constraints* are affected by his marital obligations. When time constraints exist and when energy is invested to achieve a rendezvous, positive expectations rise. Time spent together is labelled "special," "important," "encapsulated," "never mundane." Although these time constraints often become problematical later in the relationship, in the beginning they are viewed favorably (e.g., "Because our time was special, neither of us wanted to mess it up"; "There was a purity" because we "never wasted time fighting or worrying about the relationship."). "Purity" is sustained because the clandestineness of the relationship separates it from the commonplace world. The liaison does not "deteriorate" into "a day-to-day domestic situation where demands are made that are unpleasant or mundane." The couple can "stop the clock, freeze the relationship" when it is "particularly romantic . . . encapsulated."

Third, because the man is married, the single woman anticipates that the liaison will be *temporary*. In the beginning stage, not one woman expected a long-term or permanent relationship with her married lover. One, who met her lover (of four years) at a conference, stated, "I knew when the conference ended, I would never see him again and I had no intention of keeping up the relationship." Another, whose relationship lasted five years, stated, "Since I knew it was temporary—I really thought that!—I felt it was a very safe thing to do." Entering these relationships believing they will be short-lived liaisons, women can be vulnerable to their lovers

because they have no designs on his (their) future. Because the relationship is defined as temporary, it is safe. One emotionally reserved woman commented, "I'm willing to take a risk. I don't have to keep myself under control because it's not so dangerous because it can't go anywhere."

Space and time, then, are both controlled by the man's marital status. Privacy leads to semi-autonomous definitions of the meaning of the relationship, time constraints to quality time together, and temporariness to feelings of freedom and safety. This context induces feelings of great emotional intimacy.

Typically, the feeling of emotional closeness grows as the man talks about his life, disclosing his fears, hopes, and secrets. One woman said her lover told her "family secrets" that he had shared with no one else, including his wife. Another said, "I know everything about him . . . *Everything*." The woman feels *trusted*. The married man's apparent emotional openness and his willingness to be vulnerable is what draws her closer to him. In turn, the woman feels freer to talk about herself, a process much "like an add-a-pearl necklace," where both are "putting pearls [secrets] on the string." A context is established where a woman can let her "guard down" and "let a man know" more about her than "other men ever have." Through these self-disclosures, a feeling of intimacy based on "knowing about another" and "telling about oneself," as distinct from the creation of mutual secrets, is experienced.

Simmel (1950, p. 334) noted that the secret is the primary basis for individualization, and that, in modern societies, people identify more with the characteristics they hide than with the ones they publicly tout. He proposed that people tend to hide those characteristics that conflict with their public image and that they wittingly disclose secrets only to those who are indifferent to the discrepancies or who are otherwise harmless (1950, pp. 334–35). Simmel's proposal is further explicated by Davis (1973, pp. 103–4). What individuals try to hide are their *weaknesses*. People identify their uniqueness with their secret weaknesses. When people allow someone to see their vulnerabilities, they feel as though they are seen as they truly are. From Aristotle's view that "we feel friendly . . . toward those . . . who will tell us of their own weak points" (quoted in Davis 1973, p. 114), through Simmel's and Davis's percep-

tive analyses, to Lorenz's ethological research (1966), being vulnerable is arguably a way of establishing bonds. "As animals bare their throats, so humans bare their secrets" (Davis 1973, p. 115).

Once secrets about the self have been shared, Davis contends, the discloser becomes a "psychological hostage" of the listener. Knowledge confers power. If both parties exchange secrets, they reach a state of "mutual deterrence, for neither can now harm the other without having unacceptable damage inflicted upon himself in return" (1973, p. 112). Davis further proposes that "the potentially damaging secret knowledge of each other's weaknesses that both acquire" is a major force in the transformation of isolated individuals into intimates (1973, p. 130).

If this analysis is correct, the tension between concealing and revealing secrets about the self is intensified in modern societies. Sharing secrets about the self makes one a psychological hostage, but only through such sharing do people feel they are truly known as unique individuals. But does this analysis hold true for single woman/married man relationships?

Clandestine liaisons between single women and married men provide a relatively safe place for the single woman to be vulnerable. She can explore her sexuality, experiment with different ways of interacting, and expose her weaknesses (Richardson 1986). But the liaison provides a *particularly* safe place for the *married man* to be vulnerable. It is easier to reveal one's deepest feelings in a relationship that can be readily terminated and safer to discard one's veneer when there is little threat of being trapped or dominated—and easiest yet when displaying one's weaknesses is likely to enchant (Sattel 1983). Because the double standard is still operative, the risk of betrayal is minimized (Eichler 1980). The woman will be held accountable and blamed, not the man. It is his marital status that is *in fact* being protected through the secrecy. The ideology of "protecting" the woman's reputation simply reinforces his safety.

Popular belief holds that the "other woman" has the most potential power because she can destroy the man's whole life by revealing the affair. For potential power to be actual power, however, it has to be recognized and exercised. The women I interviewed did not publicly reveal their affairs or threaten to. The idea did not even occur to them. What a

woman wanted was to maintain a high-quality liaison (Richardson 1985). She had nothing to gain by threatening disclosure and a great deal to lose. Threats would diminish the quality of her relationship, if not end it. Telling the wife required acknowledging the wife's existence, which was not in the single woman's interest (Richardson 1986, pp. 88–107). If she works with her married lover, she would risk her job and be subject to accusations that she exchanged sexual favors for career advancement. Publicly, she would be labelled a "home wrecker." The fact that she did not recognize her supposed potential power to undermine his marriage testifies to its power over her. "Mutual deterrence" is not their basis for trust; minimization of exposure is. The two are not psychological hostages to each other, but trusted confidantes.

Building a relationship on an assumption of trust increases the trust, if there is no apparent betrayal. One of the consequences of increased trust is increased self-esteem. Being "needed, trusted . . . his one true confidante," the single woman feels ennobled. To be trusted with a secret means you are judged a worthy and moral person. As Simmel has commented, ". . . in the confidence of one [person] . . . in another lies as high a moral value as in the fact that the trusted person shows himself worthy of it" (1950, p. 348). Listening to his self-revelations and not betraying them, a decision which is "free and meritorious" (Simmel 1950, p. 348), can give the single woman a sense of moral rectitude and moral superiority that other elements about the relationship do not sustain (e.g., "The purity of my feelings for him were so terrific that it made me a good person"; "I still marvel that he considered me worthy enough to know him that well. His wife may have had his name, but I had his soul.").

Because the structural conditions facilitate the free exchange of secrets and an increasing positive regard for the self and the other, women understandably define their experiences as highly intimate ones. One consistent research finding is that feelings of intimacy relate to self-disclosing. Intimacy is perceived as existing when self-disclosures are freely given and are accompanied by positive feelings (e.g., Jourard 1971; Rubin 1970; and Mcallist and Bregman 1983). What the women are experiencing is the "sort of intimate exchange which we see as a primary component of love" (Rubin 1974). The

idealization of the relationship arises precisely because the man is married, with the constraints of his marital status directly affecting the construction of the nonmarital liaison. Because of his marital status, the intimates perforce become more intimate.

STAGE TWO: BECOMING A WE

If the two become trusted confidantes and establish a "special" relationship, the single woman's loving feelings may soon give way to feelings of normlessness and rootlessness, precisely because the relationship is secret. Few, if any, biographical and socialization experiences and few, if any, social supports exist to objectify and validate her relationship. But, as is true of any relationship, the subjectively experienced secret relationship has to appear objectively real in order for it to persist (Berger and Kellner 1964). In this section, I discuss the processes, including the creation of mutual secrets, through which anomie is overcome and the two become a "We."

Normally, a relationship is objectified through interaction. The married woman has the "constant conversation" with her husband to stabilize her identity, as well as the daily exchanges with friends, relatives, associates, and tradespeople "Good morning, Mrs._____.") that validate her membership in a socially accepted couple. Since the normal validations are not available to the single woman involved with a married man, the usual means for objectifying the relationship are absent, although the need is not diminished.

Simmel's (1950, p. 358) insight on ritual in secret societies is relevant here. He noted that secret societies devise rituals relatively unfettered by historical and social convention. Highly embroidered rituals can be built up from within the secret society. He proposed that the most typical feature of secret societies, the feature that most distinguishes them from open organizations, is their strong valuation of "rites, formulas, and usages." Through rituals, the members experience the secret society as a "closed unit, a whole, both sociological and subjective" (1950, p. 360).

Structures similar to those described by Simmel arise in clandestine intimate relationships, but, in addition, this research explains how the problem of objectifying the secret relationship is solved. Mutually created rites

and collected objects (mutually created se-crets) are infused with deep symbolic significance. The rites and objects become the validators. They intensify feelings and prove the relationship exists.

Every woman I interviewed volunteered an example of a rite or object that confirmed the existence of the relationship. The objects and rites signify to the woman that the relationship exists and has a history and a structure. This signification is referred to as *marking* We (Lester 1979). Some examples will clarify this process.

Some women mark We intangibly, through the use of private languages, jokes, appellations, phone codes, and celebrations of anniversaries. In describing these, the women frequently emphasize the stability of the rituals. One woman, in a long-distance relationship, commented that she and her married lover wrote each other at the same time each day, "like we were really together." Another talked about the "monthiversaries" they celebrated, and a third emphasized that Tuesday nights were routinely shared with her lover.

In addition to rites and rituals, single women mark We more tangibly, through shared possessions. One woman described in great detail how she and her lover "re-did" her bedroom—choosing colors from paint chips, repainting the walls, sanding the floors, acquiring plants and prints. She summarized, "It's something I wouldn't have done by myself or for myself. We did it for ourselves. It became *our* bedroom." Another woman reported that she began saving mementos, which she placed in a drawer. Her lover began to refer to it as "our drawer." A third said times spent looking at "their photo album" were "very special times when we could remember again." In effect, the bedroom, the drawer, and the photo album are private possessions of both the lovers in which to mark We and preserve a history.

Although probably any relationship that persists can be reconstructed by its participants in terms of their private holdings (Lester 1979), in these unsanctioned relationships, marking We carries the burden of *proof* that a relationship exists. Moreover, these private possessions remain private. Totally distinguishable from socially validated coupling in which the objects themselves (the ring, the house, the children) become the social basis for the validation, for these illicit couples, the objects are themselves secrets. Because they are not publicly shared, and yet carry so much proof of the existence of the relationship, their symbolic significance intensifies.

For women highly invested in their relationships, the objects can take on a semi-sacred quality. The reverence suggests a witnessing of holy artifacts, *enshrinement* behavior and attitudes (cf. Cook 1982). One woman, after a long introduction, brought out "their" scrapbook from its "secret hiding place." Although her manner was talkative before, while the scrapbook rested on her lap, she spoke in short descriptive statements "This is when we . . .") followed by long silences. As soon as she returned the scrapbook to its place, she resumed her earlier outgoingness. Another woman, after a long preparatory statement, entered her bedroom and positioned herself in front of a bulletin board which held a dried bouquet, a Valentine's Day Card, movie ticket stubs, and some postcards. A period of silent viewing followed. Later, she said, the board was strategically placed "so I can see it just before I turn off my night-light."

Even for women who were less highly invested in their liaisons or who had negative and angry feelings toward their lovers, the objects that signified that a relationship existed were shared only after a preamble and were handled with great respect. Because the objects are the functional alternatives to *social* confirmation and external validation, they become the manifest and material touchstones in the woman's construction of reality.

The intense meaning of the objects as representations of the relationship is further illustrated by their purposeful *destruction* to signify that the relationship no longer exists. One woman, who experienced an abusive ending to the relationship, said, "I was a wild woman. I ripped and mangled and shredded everything of ours and threw it into the trash!" Another woman, three years after her lover had walked out, said, "I spent New Year's Eve, alone, burning our pictures, one-by-one in the fireplace, chanting, 'Goodbye, good-bye,' over each one." And, a third, after mailing her lover a "Dear John" letter, placed his love letters in a box, sealed it, and "put it outside." To signify the ending of the relationship, the objects are ritually removed, the house "cleansed."

The *ending* of a socially validated relation-

ship is itself socially validated, and the processes for accomplishing it are socially prescribed. The social interaction and "constant conversation" that generated the couple's "we-ness" is replaced by social interaction that acknowledges the couple's demise. Their separation is marked linguistically, socially, and often legally. During the separating process, moreover, each of the partners can talk freely about the break-up and receive social support.

The same autonomy and freedom from social norms that affected the construction of the secret liaison affect its ending. Neither social interaction nor public rituals are available to signify to the self that the relationship no longer exists. Because the relationship is secret, there are few (if any) significant others to talk to about it and even fewer who will provide emotional support. Consequently, the single woman has a paucity of forms through which to "deconstruct" her relationship. The one way she does have is the destruction of the objects that signified its existence. The willful annihilation of the same objects that validated the existence of the relationship signify its demise.

But before the relationship ends, because these rituals and objects are mutually created secrets, another kind of intimacy is experienced. This is intimacy arising from feeling *attached to* another person, *a part of* a relationship, a whole, both objectively and subjectively. This feeling, like the intimacy arising from exchanging secrets, is one associated with being in love (Rubin 1974). The single woman, consequently, exists in a social reality of a secretly shared world.

Existing in that world, however, has differential consequences for her than for her married lover precisely because she has deferred to his marital status and allowed it to determine the contours of her liaison. The more objectively real and sacrosanct the relationship to her, the more she comes to value the liaison. Consequently, she becomes more committed to preserving it. The more dedicated she is, the more disempowered she becomes. The disempowerment is intensified because of the man's marital status. First, she is denied the social prestige attached to "having" a man and "being in love." Her ability to "attract" a man is not publicly acknowledged. As a result, she becomes more, rather than less, dependent on the

married man's attentions. But, even more disempowering is how she constructs her life to protect the secret of the relationship's existence. She deepens her sense of intimacy and commitment because she decreases the opportunities for publicly testing the relationship and/or increases her dependence on it.

Because the relationship is a secret one, it is not socially tested. She can cherish her idealization because the relationship is insulated from external judgments and critical appraisals. But, in addition, because she *wants* to see him, she may find herself, "arranging [her] whole life around his, deciding when . . . [she] can go out and when . . . [she] should stay home." To reduce the chances of revealing her secret, she may withdraw from her usual social network and from her family and friends, which makes her even less socially moored and her social identities more fragile. One woman commented, "In a normal relationship you usually don't have to give up who you are. In my relationship, I gave up my family, my identity, my culture, the theater, the arts . . . my therapist." Conversations with friends and family may be truncated as the single woman becomes more matter-of-fact and distant in her intentions, revealing little about her life for fear of revealing all (e.g., "I just repressed my life. I didn't talk about anything to anyone."). If she retreats further into social and emotional isolation, her married lover grows more salient. He becomes *de facto* her "whole life." Her withdrawal from others, then, intensifies her dependence on the relationship, her commitment to it, and her love feelings. He is her intimate.

Not all women withdraw so totally from friends and family. However, all the women I interviewed reported episodes such as "lying on the couch on a Sunday afternoon drinking gin, crying, and waiting for him to call"; behavior changes such as "seeing much less of my female friends, so I'd be home if he should call"; routine changes such as "dropping out of the car pool so I could drive the long way past his office"; distancing activities such as "lying to my friends and family about who I was seeing and what I was doing with my time"; and compartmentalizing their liaison, living a disjointed, rather than an integrated life. Equivalent socioeconomic statuses do not prevent women from relationship disempowerment. Indeed, these

illustrations are from women who were socioeconomic equals to their married lovers.

To some degree, women who manage to sustain the relationship sacrifice some independence and autonomy as they construct their lives to accommodate the man's marital status. Their gender contributes to their willingness to (and skill at) accommodating the male, but, it should be reemphasized, it is primarily because of his marital status, not his gender, that she reconstructs her life in the way she does. Unlike these clandestine liaisons, normative male-female relationships are openly constructed, unfettered by a man's commitment to another woman, and have benchmarks that signal the relationship is progressing toward permanency or termination. And they do not require the woman to keep her feelings and significant holdings a secret from others.

The greater the woman's interest in the relationship, the more disempowered she becomes. The less her interest, the more power she retains. The acts required to conceal the relationship create new secrets, limit external testing of the relationship, increase her dependency, and reduce her opportunity for alternative relationships. Understandably, her feelings of intimacy intensify while her power within the relationship diminishes. As a final note, even her power to *end* the relationship diminishes, the more invested in it she becomes. The more dependent the woman on the relationship, the more likely her lover will decide when and how to terminate the relationship, leaving her to emotionally free herself from the liaison, as well as from her feelings of being rejected or abandoned (Richardson 1979).

DISCUSSION

This paper has gone beyond Simmel by looking at how status and secrecy affect the social construction of secret liaisons between single women and married men. Several salient elements have been identified: how the suspension of the normal rules allows the construction of an alternate reality; how, because the relationship is forbidden, it is not socially tested, thereby contributing to its longevity; how anomie arises and is managed through imbuing rites and objects with intense symbolic significance; how the man's marital status is structurally conducive to idealization of the relationship; and how prioritizing the

man's marital status disempowers the woman. These elements may be the same ones that characterize other kinds of forbidden sexual relationships. Although the particulars of any given forbidden relationship may be unique, the contours of them as a *category* may have similar characteristics.

If a status of one of the parties in a secret, sexual liaison is given priority, the relationship will be constructed to protect that person. Subsequently, that person will have greater power in the relationship and greater control over the relationship's construction and denouement. If no status of either party is given priority, the relationship is more likely to be power-balanced. A primary example of this is an extramarital relationship between two married people. Because both parties are in equal need of protection and because both parties have the power that accrues from being married, the imbalances are minimized (Atwater 1982; Lawson, forthcoming).

Marriage, I have argued, confers upon the married greater prestige, greater social-psychological and structural support, and, consequently, greater relationship power than singlehood confers upon the single. Although the affects of marital status are subtle and often masked by more glaring status differences such as gender and socioeconomic status, interactional consequences are real and significant. The status of being married grants an interactional edge to those who are married and elicits an interactional deference to the status from those who are not.

If being married confers as much interactional power as I propose it does, married women involved with single men should enjoy greater power in their liaisons than single women do with married men. At-water's (1982; 1984) research on "the extra-marital connection" of married women supports this claim. Married women enjoy considerable power in their relationships with single men—much more than single women do with married men. The status, married, enhances the gender status, female. The structural limitations of being married work for her, not against her. She has an "oversupply" of men, she is socially moored in a marriage, she can negotiate "emotional equality" she feels in control of the relationship, she is more likely to determine when and how to terminate the relationship than he is, and she has a social structural base to

return to when the relationship ends (Atwater 1984).

This is not to argue, however, that marital status completely overrides gender. We would expect an interplay between gender and marital status. For example, it is unlikely that a single man involved with a married woman would act submissively, put his life on hold, or do the heavy emotional work necessary to sustain the relationship. His gender, male, will have consequences for his experiences in liaison construction that further research would specify.

The research on single women/married men liaisons challenges the implicit assumption that gender is invariably the primary factor in the construction of male-female relationships. The research is consonant with the research agenda of contemporary gender sociology, to discover how gender is shaped in concrete social relationships and to determine its relative effects in different situations (Thorne, Kramarae, and Henley 1983; Richardson 1988; West and Zimmerman 1987). The finding points to an area of analysis that has been overlooked by both gender and family sociologists, namely, "marital privilege" as a determinant and/or mediator of interactional power.

Finally, because we live in a complex, gesellschaft society, it can be argued that the need for secret, forbidden relationships has increased. Secondary relationships, homogenization, and instrumentality reduce the opportunities for individuation and personal trust. Secret relationships provide a harbor from the normative world where an apparently ideal relationship can be constructed and individuation and trust created.

But for whom is the harbor safe? Secret relationships protect the interests of those with the greater status and power—the married man (woman) rather than the single woman (man), the doctor rather than the patient, the adult rather than the child. The more impermeable the boundaries and the greater the status differential, the more likely the interests of the powerful will be met through the construction of secret relationships. Status differences are carried into the forbidden liaison, and rules are generated that protect the person with higher status. In a very profound sense, then, secret, forbidden, sexual relationships are no sociological surprise. They reinforce and perpetuate the interests of the powerful.

REFERENCES

Atwater, Lynn. 1982. *The Extramarital Connection: Sex, Intimacy, and Identity*. New York: Irvington.
———. 1984. "Gender and Norms in Extramarital Relationships." Paper presented to the National Council of Family Relations, San Francisco.
Berger, Peter and Brigette Berger. 1975. *Sociology: A Biographical Approach*. New York: Basic Books.
Berger, Peter and Helena Kellner. 1964. "Marriage and the Construction of Reality." *Diogenes* 46:1–25.
Blau, Peter M. 1964. *Exchange and Power in Social Life*. 1964. New York: Wiley.
Cook, Judith A. 1982. "The Adjustment of Parents Following the Death of a Child from a Terminal Illness." Ph.D. diss. Ohio State University.
Davis, Murray S. 1973. *Intimate Relations*. New York: Free Press.
De Rougemont, Denis. 1956. *Love in the Western World*. Princeton, NJ: Princeton University Press.
Diamond, Timothy. 1983. "Caring Work." *Contemporary Sociology* 13:556–58.
Eichler, Margrit. 1980. *The Double Standard: A Feminist Critique of Feminist Social Science*. New York: St. Martin's Press.
Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA.: Harvard University Press.
Glaser, Barney G. and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
Hochschild, Arlie Russell. 1983. *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.
Jourard, Sidney. 1971. *Self-Disclosure: An Experimental Analysis of the Transparent Self*. New York: Wiley-Interscience.
Karp, David A. and William C. Yoels. 1979. *Symbols, Selves, and Society*. New York: Harper and Row.
Koch, Michael. 1980. "Sexual Abuse in Children." *Adolescence* 15:643–48.
Lawson, Annette. Forthcoming. *Adultery*. New York: Basic Books.
Lester, Marilyn. 1979. "Making Music Together: A Sociological Formulation of Intimate Encounters between Males and Females." Presented at the American Sociological Association Meetings, Boston.
Lorenz, Konrad. 1966. *On Aggression*. New York: Harcourt, Brace and World.
Mcallist, Hunter A. and Norman J. Bregman. 1983. "Self-Disclosure and Liking: An Integration-Theory Approach." *Journal of Personality* 51:202–7.
MacKinnon, Catherine. 1979. *Sexual Harassment of Working Women*. New Haven, CT: Yale University Press.
Reiss, Ira, R.E. Anderson, and G.C. Sponaugle. 1980. "A Multivariate Model of Determinants of Extramarital Permissiveness." *Journal of Marriage and the Family* 42:395–411.
Reskin, Barbara and Heidi I. Hartmann. 1986. *Women's Work and Men's Work: Sex Segregation on the Job*. Washington, DC: National Academy Press.
Larry T. Reynolds. 1987. *Interactionism: Exposition and Critique*. Dix Hills, NY: General Hall.
Richardson, Laurel [Walum]. 1979. "The 'Other Woman': The End of the Long Affair." *Alternative Lifestyles* 2:397–414.
———. 1985. *The New Other Woman: Contemporary*

Single Women in Affairs with Married Men. New York: Free Press.

———. 1986. "Another World." *Psychology Today.* February: pp. 22–27.

———. 1988. *The Dynamics of Sex and Gender: A Sociological Perspective,* third Rev. Ed. New York: Harper and Row.

Rubin, Zick. 1970. "Measurement of Romantic Love." *Journal of Personality and Social Psychology* 16:265–73.

———. 1974. "Lovers and Other Strangers: The Development of Intimacy in Encounters and Relationships." *American Scientist* 62:182–90.

Russell, Diana E.H. 1986. *The Secret Trauma: Incest in the Lives of Girls and Women.* New York: Basic Books.

Sattell, Jack W. 1983. "Men, Inexpressiveness, and Power." Pp. 242–46 in *Feminist Frontiers,* edited by Laurel Richardson and Verta Taylor. New York: Random House.

Simmel, Georg. [1902–03; 1908; 1917] 1950. *The Sociology of Georg Simmel,* Translated by Kurt H. Wolfe. New York: Free Press.

Stein, Peter J. 1981. "Understanding Single Adulthood." Pp. 9–21 in *Single Life: Unmarried Adults in Social Context,* edited by Peter J. Stein. New York: St. Martin's.

Thorne, Barrie, Cheris Kramarae, and Nancy Henley. "Language, Gender, and Society: Opening a Second Decade of Research." Pp. 7–24 in *Language, Gender and Society,* edited by Barrie Thorne, Cheris Kramarae, and Nancy Henley. Rowley, MA: Newbury House.

Waller, Willard. 1938. *The Family: A Dynamic Interpretation.* Hinsdale, IL: Dryden Press.

West, Candace and Don H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 2:125–51.

POWER RELATIONS IN EXCHANGE NETWORKS*

BARRY MARKOVSKY   DAVID WILLER
University of Iowa   University of Kansas

TRAVIS PATTON
University of Kansas

Many theories address the problem of how a social structure affects the experiences and behaviors of its members. This paper offers a network-exchange theory to solve this problem. Previous research has shown that the nature and outcomes of negotiations among individual or corporate actors can be inferred from their network positions. The impact of this research has been limited because its theory does not enable the researcher to locate power positions in the networks. We offer a theory that is both consistent with all previously reported experimental research and is generalized to conditions not considered by other formulations. In addition to supporting derived hypotheses pertaining to network-based power, our experiments demonstrate, among other things, that certain unstable networks break down to form stable substructures and that some networks contain overlapping but autonomous domains of power and exchange.

Although no single exchange theory dominates the social sciences, a fairly coherent social-exchange perspective exists. In this perspective, social structures and processes impinge on and emerge from resource and sanction transfers between individuals and/or collectivities.[1] Recently, some theories have moved beyond two-party exchange contexts to focus on networks of exchange relations. As structural theories, network-exchange theories attempt to explain how macro-properties bear upon micro-units within structures. Concretely, they try to show how network structures affect the power of actors to extract valued resources in their exchanges with others.

We propose and test a theory that predicts relative power for network positions. In so doing, we address several structural phenomena, including the breakdown of larger networks into smaller parts and the emergence of positions that simultaneously have one level of power in one part of the network and a different level in another. Our theory is further intended to provide higher levels of rigor, power, and specificity than are found in earlier approaches. We find that each such technical advancement produces a manifold increase in the array of potential applications.

Whenever a person or group negotiates with another person or group over the allocation of valued resources, a minimal social-exchange network exists. More elaborate (i.e., nondyadic) structures form when one member is involved in two or more such relations. For example, college students Al, Bea, and Cleo each want to date, and norms prohibit them from dating more than one person at a time. Suppose that Bea and Cleo both vie for Al's attention and have no other prospects, while Al would be happy to date either Bea or Cleo. This creates a B-A-C network, where A(1) may "negotiate" with B(ea) and C(leo), but only date one of them. Such circumstances actually do tip the balance of power (Peplau 1979) in dating relations: A is able to make greater demands than his chosen partner, and generally has greater influence in the relationship. But if B or C develop dating interests with a responsive D, A loses his structural advantage.

[1] Theoretical statements have been provided by Thibaut and Kelley (1959), Blau (1964), Gergen (1969), Homans (1974), Ekeh (1974), Heath (1976), Blalock and Wilken (1979), Burgess and Huston (1979), and Cook (1987). Emerson (1976), Bredemeier (1978), and Tuner (1986) have written reviews. Applications involving ethnographic, institutional, and historical analyses are provided by Polanyi (1944), Elkin (1953), Sahlins (1972), Earle and Ericson (1977), and Emerson (1981). Recent applications of *network*-exchange theories to interorganizational relations, backward and forward integration of the firm, community structure, historical development of modern exchange relations, and exchange processes in antiquity are given by Hansen (1981), Loukinen (1981), Gilham (1981), Galaskiewicz (1985), Skinner and Guiltinan (1986), Lind (1987), and Willer (1987).

This type of analysis is applicable in other areas such as international, auctioneer-bidder, retailer-consumer, and manufacturer-retailer relations. A good example is the control that a manufacturer may impose upon retailer marketing strategies (Skinner and Guiltinan 1986). Suppose Ascii Ugetty (A) is the sole manufacturer of a line of computer games. Big Bytes (B), Chips-R-Down (C) and Data Dump (D) are independent retailers that want to carry the line. Even with fixed wholesale prices, A's position affords it power over B, C, and D. Skinner and Guiltinan found that retailer activities such as advertising expenditures, sales force training, and credit policies were under manufacturer control to a greater extent when retailers had no alternative suppliers. So if D can obtain the product from E–Z Access (E), A loses its ability to control D's policies. A may have to "outbid" E just to keep D's business.

Our purpose is to understand the structural logic manifested in all such exchange networks—a logic unbounded by empirical content. If the experiences of actors depend on their positions, this suggests a structural determination of behavior. At issue in this paper is the logic of that determination.

AN EARLIER APPROACH

Recent work by Cook, Emerson, Gillmore, and Yamagishi (1983) clearly overlaps with our own in scope.[2] They showed that their approach could anticipate power distributions in some cases where alternative measures failed. Based on Emerson (1972b), Cook et al. (1983) defined Exchange network as

(1) a set of actors (either natural persons or corporate groups), (2) a distribution of valued resources among those actors, (3) for each actor a set of exchange opportunities with other actors in the network, (4) a set of historically developed and utilized exchange opportunities called exchange relations, and (5) a set of network connections linking exchange relations into a single network structure. (p. 277)

The set of exchange relations is a subset of exchange opportunities, and actors in the

system are assumed to be committed to exchanging within their relations, to the exclusion of alternative opportunities. The concept of *connection* permits networks to be considered from relations. Formally

Two exchange relations between actors A-B and actors A-C are connected to form the minimal network B-A-C to the degree that exchange in one relation is contingent on exchange (or nonexchange) in the other relation. (a) The connection is positive if exchange in one relation is contingent on exchange in the other. (b) The connection is negative if exchange in one relation is contingent on nonexchange in the other. (p. 277)

A negative connection exists if B and C can substitute as providers of A's resources. The authors cite as examples dating and friendship networks. In the case of a positive connection, A cannot benefit without exchanges from B *and* C. This is true if A is a brokerage agent, or if B and C are assembly-line workers who must exchange their labor for pay before the firm (A) can benefit.

Cook et al. (1983 define *power* as "In any dyadic exchange relation $A_x;B_y$ (where A and B are actors, and x and y are resources introduced in exchange), the power of A over B is the potential of A to obtain favorable outcomes at B's expense" (p. 284). *Dependence* is given as: "The dependence of A on B in a dyadic exchange relation is a joint function (1) varying directly with the value of y to A, and (2) varying inversely with the availability of y to A from alternate sources" (pp. 284–85).

By *informally* applying power-dependence ideas, Cook et al. developed several hypotheses predicting relative power for positions in several types of negatively connected networks. Toward the end of the paper, a network *vulnerability* (V) method was suggested as a first step toward a formal procedure for predicting positions' relative power.

To determine V for the B-A-C network, assume that related actors negotiate over the division of 24 resource points, and a one-exchange rule creates the negative connection: A may exchange with B or C but not both in a given round. First, the maximum resource flow (MRF) for the network is calculated. MRF = 24 since, by the one-exchange rule, only 24 points may be distributed per round. Next, the reduction in

---

[2] Comparisons among these theories are hindered by their lack of explicit scope conditions. Although some scope conditions can be inferred, at times it is not clear when theories are competitors (Wagner and Berger, 1985) with divergent predictions testable in the same empirical settings.
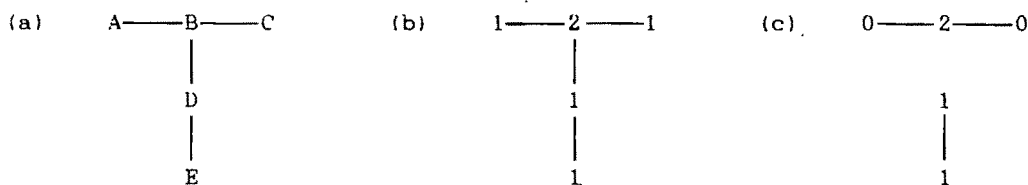
Fig. 1

maximum flow (RMF) is calculated for each position by noting the effect of its removal on MRF. If B or C is removed, RMF = 0, since A may still exchange with the other. However, $RMF_A = 24$. The network is then most vulnerable at A, and A is declared a power node.

Discussion

This general approach has been corroborated in several experiments, including those published by Cook et al. (1983), Stolte and Emerson (1977), and Cook and Emerson (1978). However, V has not been systematically tested. Moreover, Willer (1986) determined that V produces untenable predictions for some relatively simple networks such as that in Figure 1a. V predicts high power for B, D, and E. But under Cook et al.'s experimental conditions, high profit for D would entail low profit for B, E, or both. Although Cook, Gillmore, and Yamagishi (1986) described V as only "a preliminary notion," it still provided the only explicit basis for deriving hypotheses. Without it, predictions were informal and not fully determined by the theory.

Cook et al. (1986, p. 447) later proposed a modified V-measure. Network-wide dependence (D_N) weighs a position's RMF by the factor $(1 - CRMF)$, where CRMF is "no. of lines that need to be removed [for a position] to exercise power at its potential" divided by the number of lines connected to the position. By this measure, B in Figure 1 has higher power than D, and D higher power than A, C, and E. Although these predictions are tenable, they diverge from test results reported later in this paper and their derivation is indeterminate.[3]

Many of Cook et al.'s methodological choices were neither necessitated nor precluded by their theory. For instance, negotiations took place over a series of rounds; each relation had its own resource pool; each pool was replenished before every round; exchange consisted of mutually agreed on pool divisions; there was a one-exchange rule; resources did not move through positions; coalitions were prohibited; and actors had no information on negotiations in which they were not directly involved. At issue is whether the approach might have been falsified under alternative methodological conditions. Later we demonstrate that very different results are obtained under slightly different conditions.

A GRAPH-ANALYTIC THEORY

In his recent elaboration on the work of Cook and her associates, Marsden (1987) succinctly offered as unsolved problems several of the implications that may be drawn from our theory:

> The difficulty in developing a more general measure is that an alternative [exchange partner] may be exploitable for two reasons: It may have few alternative relations, or all of its alternatives (irrespective of how many in all are available) may be in a position to exploit others. The second condition of exploitability can lead to consideration of quite distal features of network structure. (p. 147, note 5)

Building on an earlier exchange formulation (Willer and Anderson 1981; Willer 1987), our graph-analytic approach recognizes both types of "exploitability" and

---

[3] Using their model, we could not reproduce Cook et al.'s predictions. The authors stated "This measure is relevant only when RMF is not zero" (p. 447). But RMF = 0 for positions A and C, apparently making D_N inapplicable. Further, the expression "exercise power at its potential" is not defined, and it is not stated whether

the removed lines must stem from the position whose D_N is being assessed. Following the Cook et al. examples, it appears that in the 1a network, two lines must be removed from B to reduce the maximum flow of network resources, and one relation must be removed from D. The result is $CRMF_B = \frac{1}{3}$, $CRMF_D = \frac{1}{2}$, $D_{NB} = 8$, and $D_{ND} = 12$. D should be higher than B, contradicting Cook et al.'s prediction. In either case, the predictions diverge from those we will obtain from our model.

specifies conditions under which distal network properties will or will not influence proximal outcomes. We first present $p(1)$, an index for power in one-exchange networks. This allows us to test our predictions against those of Cook et al. (1983, 1986). Following this, $p(e)$, a generalized version, will be explicated and tested.

Conditions of Exchange

Power and resource distributions are affected not only by network shapes, but also by the conditions under which exchanges transpire. The theory provides scope statements encompassing relatively broad conditions, some of which are later relaxed, others of which await future tests, theoretical extensions, and refinements. Scope conditions are not assumptions about human nature or frequencies of empirical circumstances. They are statements that, if satisfied (or approximated), commit the theory to critical examination and, if not satisfied, relieve it of any explanatory imperative (Walker and Cohen 1985).

Several important concepts must first be defined: *actors* are decision-making entities, e.g., organisms, collectivities, or even computer programs. *Positions* are network locations occupied by actors. A *relation* between two positions is an exchange opportunity for actors in those positions. In short, *actors* occupy *positions* linked by *relations*.[4] We will index both actors and positions using upper:case letters and at times refer to them interchangeably.

Actor Conditions. Four conditions delimit actors' behavior: (1) all actors use identical strategies in negotiating exchanges; (2) actors consistently excluded from exchanges raise their offers; (3) those consistently included in exchanges lower their offers; (4) actors accept the best offer they receive, and choose randomly in deciding among tied best offers.

Condition 1, requiring identical strategies, is nearly always implicit in exchange theories. In tests and applications, however, it is generally sufficient that actors adopt functionally *similar* strategies. Condition 1 also asserts that actors negotiate, i.e., they make offers and adjust their subsequent offers in

light of counter-offers they receive. Conditions 2 and 3 require that actors seek to enter exchange if previously denied, and to improve outcomes beyond those previously obtained. Finally, condition 4 rules out a range of strategies that may drive up the offers of excluded parties.[5]

Position Conditions. These apply to positions and their relations: (5) each position is related to, and seeks exchange with, one or more other positions; (6) at the start of an exchange round, equal pools of positively valued resource units are available in every relation; (7) two positions receive resources from their common pool if and only if they exchange; (8) each position exchanges with at most one other position per round.

Since isolates cannot exchange, Condition 5 omits them from consideration. Condition 6 reflects conditions in most prior research: a pool of *profit points* resides in every relation and is replenished with each new round. Condition 7 indicates that two actors will not exchange unless both benefit. Condition 8, relaxed later, asserts that actors may complete at most one exchange per round. This creates negative connections in a way consistent with all previously cited experimental research and Cook et al.'s (1983) simulations. It assumes that, for whatever reasons, actors only require a single exchange, or are only able or permitted to complete a single exchange in a given round.[6]

The Graph-theoretic Power Index

Building upon simple arithmetic procedures, our graph-theoretic power index (GPI) deter-

---

[4] The reason for distinguishing actors and positions is that actor properties (e.g., decision strategies) and position properties (e.g., number of relations) may affect power independently (Markovsky 1987a).

[5] These conditions allow a variety of more determinate rational or quasi-rational strategies. For example, resistance theory (Heckathorn 1980; Willer 1981, 1987) provides an elegant model of joint-bargaining decision-making. Resistance is given as the ratio of an actor's interest in gaining a better exchange to interest in avoiding conflict. The conditions do, however, rule out strategies such as coalition formation (Kahan and Rapoport 1984; Shubik 1982; Willer 1987), in which some actors temporarily accept reduced resources while receiving increasingly favorable offers from others.

[6] We treat negative connection the same way as Cook et al. (1983), but diverge from Emerson's (1972a,b) original usage (Willer, Markovsky, and Patton forthcoming). In the earlier formulation, for an actor with multiple relations, exchange in one reduces the value of exchange in others because the actor's satiation level increases with each exchange. Exchange rates across the actor's relations are then negatively correlated, but as an *outcome* of the exchange process, not as an *initial condition*.

F₁——— E₁——— D———E₂———F₂
                    |
                    |
                    E₃
                    |
                    |
                    F₃

Fig. 2

mines relative power for all positions in any network that meets the scope conditions.[7] As also implied in the work of Kuhn (1974), Cook et al. (1983), Bonacich (1987a), Marsden (1983, 1987) and others, power is assumed to derive from the availability of alternative exchange relations, the unavailability of their relations' alternative relations, and so on. *Power* is then conceived as an unobservable, structurally determined potential for obtaining relatively favorable resource levels. *Power use,* as manifested in resource distributions, serves as an indicator of power. So while we theorize about potential power, we test our theory by observing its use.

The procedure for determining GPI involves counting path lengths. Thus, network B-A-C has two one-paths, A-B and A-C. B and C are linked by a two-path. As explained below, path counting is greatly simplified by only counting the number of *nonintersecting* paths of each length stemming from a given position. Nonintersecting paths stemming from position X have only X in common. In Figure 2, for example, three nonintersecting two-paths stem from D, but only one nonintersecting two-path stems from E₁ (connecting with either E₂ or E₃).

An implication of this procedure is that it does not matter for X whether a position *m* steps away "branches" to one or a hundred positions *m* + 1 steps away. All that matters is *whether or not* there is a position *m* + 1 steps from X. This is a subtle, possibly nonintuitive, but incontrovertible assertion within our framework. The following example therefore bears careful study.

Imagine removing A and C from the Figure 1a network. D benefits greatly from the resulting three-actor chain: B and E must try to engage D, offering ever more favorable deals to D. *Now restore A.* B now has an alternative to bidding against E. But with B not bidding against E, D's advantage dis-

solves. Although still with two alternatives, D cannot play B and E against each other and so all positions are on an equal footing. *Now restore C.* B now benefits because A and C will try to outbid each other for B's exchange. This presents no further disadvantage for D, however, who may still exchange with E on an equal basis.

Note that A and C are on intersecting two-paths from D. The creation of one of those two-paths changed the minimum relative power in D's relations from high to equal. *But the creation of the second two-path had no effect on this minimum.* If we further attached F, G, and H to position B, these added two-paths from D will still not affect the minimum relative power that D would enjoy. This shows why only one nonintersecting path of a given length is counted.

It may now be apparent that X's odd-length nonintersecting paths are advantageous, and even-length nonintersecting paths are disadvantageous. *Advantageous paths* either provide direct exchange alternatives (in the case of one-paths), or counteract the advantage-robbing effects of *disadvantageous paths.*

The GPI simply tallies the number of advantageous paths and subtracts the number of disadvantageous paths to determine each position's potential power.

Position i's GPI under the one-exchange condition is calculated as[8]

$$p(1)_i = \sum_{k=1}^{g} (-1)^{(k-1)} m_{ik}$$
$$= m_{i1} - m_{i2} + m_{i3}$$
$$- m_{i4} + \ldots \pm m_{ig} \quad (1)$$

and i's power relative to j is

$$p(1)_{ij} = p(1)_i - p(1)_j.$$

The function $(-1)^{(k-1)}$ produces + signs for advantageous paths and − signs for disadvantageous paths. These are attached to the m_{ik} values—the number of position i's nonintersecting paths of length k. For now we may

---

[7] See Harary, Norman, and Cartwright (1965), Harary (1969), and Fararo (1973) for discussions of a variety of graph-theoretic tools.

---

[8] Readers familiar with our unpublished reports should note that we have referred to this measure as $C_N(i)$, position i's centrality when allowed N exchanges. The present notation more accurately reflects our concern with power rather than centrality and adheres to the convention of displaying variable indices and parameters as, respectively, subscripts and parenthetical elements.

suppress the number-of-exchanges parameter for $p(1)_i$ and refer to the index simply as p_i.

The values for g and the ms are obtained as follows:

> m_{i1} is the number of one-paths stemming from position i, which is the same as the number of i's relations. In Figure 1a, for example, $m_{D1} = 2$.
>
> m_{i2} is the number of nonintersecting two-paths from i. As shown in the earlier example, D has only one nonintersecting two-path, so $m_{D2} = 1$.
>
> m_{i3} is the number of nonintersecting three-paths stemming from i; $m_{D3} = 0$.

The largest path of length k for which $m_{ik} > 0$ is the *geodesic* (g) of the network. In Figure 1a, three-paths link A to E, and C to E, hence, $g = 3$.

The final step is to combine the m_{ik}s: take m_{i1}, subtract m_{i2}, add m_{i3}, and so on. We find that $p_D = 2 - 1 + 0 = 1$. Figure 1b shows this value and the p_i values for the other four positions.[9]

Axioms and Theorems

The formal statement of our theory appears in the Appendix to this paper. In the statements below, "power" refers to p_{ij}, with i and j related.

> AXIOM 1: given by equation (1) above.
>
> AXIOM 2: i seeks exchange with j if and only if i's power is greater than j's, or if i's power relative to j equals or exceeds that in any of i's other relations.
>
> AXIOM 3: i and j can exchange only if each seeks exchange with the other.
>
> AXIOM 4. if i and j exchange, then i receives more resources than j if and only if i has more power than j.

In Axiom 2, "i seeks exchange with j" means that i makes competitive offers to j, i.e., offers that compete with others that j receives. A more psychological interpretation would be "i makes offers that j seriously considers." The axiom first claims that this occurs if i's power is greater than j's. Further, even if i's power is less than j's, i will seek exchange with j if i's relative power is even lower in its other relations.[10] Note that Axiom 3 *does not* imply that two actors will exchange if they seek exchange with each other; actors may negotiate without exchanging. Finally, Axiom 4 asserts that potential power determines the use of power, i.e.,GPI predicts final resource distributions.

Some of the theorems that can be derived from these axioms include

> Theorem 1: If i has no alternative relations, then i seeks exchange with j.
>
> Theorem 2: If i does not seek exchange with j *or* if j does not seek exchange with i, then i and j do not exchange.
>
> Theorem 3: Actor i does not seek exchange with j if and only if i's power is less than or equal to j's *and* i has a better alternative to j.
>
> Theorem 4: If i's power is less than or equal to j's and i has a better alternative to j, or if j's power is less than or equal to i's and j has a better alternative to i, then i and j will not exchange.

More intuitively, Theorem 1 claims that an actor in a position with only one relation will seek exchange via that relation, whatever its relative power. Theorem 2 is a logical variant of Axiom 3. Theorem 3 specifies the conditions under which an actor will not seek exchange via one of its relations. Theorem 4 predicts when a network will break at the i-j relation. It reveals that certain relations are expected to remain unused, leading some

---

[9] Exchange in one relation will often temporarily alter the relative power of nearby positions. This dynamic is captured through an iterative application of the GPI. In Figure 2, for example, p_i is first calculated for all positions in the network. In a given round of negotiation, if E_1 and F_1 exchange first, p_i is recalculated for the network with E_1 and F_1 removed. The new p_i values are then in force until the next exchange occurs or until the end of the round. In the relatively simple networks examined in this paper, initial p_i values provide accurate predictions for power use. In more complex networks, however, the iterative application of the GPI is required to obtain accurate predictions (Markovsky, Willer, and Patton 1987).

[10] After a sufficiently extended series of exchanges, an actor with $p = 0$ should seek exchange in *all* of its relations, regardless of power differences. That is, to avoid complete exclusion, the actor will offer to keep just one resource unit and relinquish the balance of the pool to any other that is willing to exchange. This seems to violate Axiom 2; however, this actor is no longer engaged in negotiation. This violates the first actor condition and makes the theory inapplicable. This is hardly a limitation of the theory, however, for when exchanges reach this point of non-negotiability, the system (or subsystem) has run its course, exchange rates will remain fixed, and the theory is "finished" with its predictions for the application.

complex networks to break apart into smaller, stable subnetworks. When such a break occurs, power indices are recalculated within the resulting subnetworks. This is demonstrated in some of the applications below.

Applications

We have applied the GPI, axioms, and theorems to a large number of networks of varying shape and size. This small sampling demonstrates the use of the theory.

For the A-B dyad, $p_A = p_B = 1$. No position has a structural advantage. The same is true for positions on any even-length chain, as verified in computer simulation research (Markovsky 1987b). In general, however, the longer the chain, the more rounds transpire before the predicted power relations stabilize.

For the B-A-C network, $p_B = p_C = 1 - 1 = 0$ and $p_A = 2 - 0 = 2$. A's power advantage is 2 in both of its relations, while $p_{BA} = p_{CA} = -2$. In fact, for odd-length chains of any length, $p = 2$ for even positions and $p = 0$ for odd positions; low- and high-power positions alternate. This conforms with Cook et al.'s predictions and experimental results for the five-position chain and with our computer simulations for longer chains.[11] Similarly, in Figure 2, $p_F = 1 - 1 + 1 - 1 = 0$, $p_E = 2 - 1 + 1 = 2$, and $p_D = 3 - 3 = 0$. Thus the center and periphery have low power, and the off-center positions have high power. This also conforms with Cook et al.'s predictions and simulation results.

Returning to Figure 1, we find that a decomposition is predicted. Figure 1b shows the p_i values as initially calculated. Applying Theorem 4, however, since D's index is less than B's, and since E is a "better" alternative for D (because $p_E < p_B$), D and B are predicted not to exchange. Finally, Figure 1c shows the final p_i values recalculated for the resulting subnetworks.

EXPERIMENT 1

Since the scope of our theory appears to overlap with that of Cook et al., we compare our predictions with those derived from their measure. We tested the Figure 1 network.

Based on our analysis, D-E will form an equal power dyad, the B-D relation will break, and B will have power over A and C. In contrast, Cook et al. (1986) order B > D > (A, C, E) with no breaks predicted.

Method

Subjects were undergraduates at a large university. Before being taken to the laboratory, participants in a given session met as a group, received written instructions, and had any questions answered. In the research room, connections among network positions were clearly marked and, to limit collusion, temporary barriers separated positions among which exchange was prohibited. The setting minimally restricted the availability of information about the structure and the actions of others.[12]

Twenty-four counters were placed between related positions. These served as resources to be divided by mutual agreement, each valued at one profit point and worth 3 cents. Each position was limited to one agreement per round. Before starting, we emphasized that exchanges could only occur by mutual agreement between related positions, and long-term strategies were prohibited.

Experiments were organized by rounds, periods, and sessions. In all, five sessions were run, each with a different group of subjects. There were five periods per session, allowing each subject to occupy each position for one period before the session was over. Each period contained four negotiation rounds, each with a three-minute time limit. Each position's scores were announced after every round. At the close of a session, participants were paid according to points they obtained— around $5.00 on the average. This design produced a total of 100 rounds of negotiation.

Hypotheses

Below we present hypotheses derived from our theory, those obtained from Cook et al.'s (1986) D_N procedure, and the null hypotheses.

1. Our theory predicts that the network will break at the B-D relation, eliminating ex-

---

[11] Cook et al. had a "low profit" relation between the two end-points of the chain. While this places a lower limit on the profit that these positions can receive, it does not affect the relative power of positions in this network.

[12] Having information on negotiations other than one's own is expected to accelerate the use of power, but not affect relative power. For a more extended discussion of information effects, see Willer and Markovsky (1986).

Table 1. First Experiment: Profit by Position

Session	Position					Test	
	E	D	A	B	C	t^*	p
1	12.55	11.42	4.29	19.10	5.09	7.85	<.0005
2	12.45	11.58	8.56	15.33	8.25	2.01	<.025
3	12.00	12.00	3.29	20.95	3.29	3.50	<.0005
4	12.05	11.95	3.75	21.55	3.75	11.15	<.0005
5	11.80	12.20	4.17	19.16	4.17	5.77	<.0005

* The reported tests are for position B's actual profit points versus the null hypothesis of 12 profit points.

change between B and D. D_N provides no hypothesis in this regard. In contrast, if exchanges are distributed randomly in the network, B will turn to D one-third of the time, but half of those times D will turn to E. The null hypothesis, then, predicts .333 × .500 × 100 = 16.667 exchanges between B and D.

2. B will exercise power over A and C, so B will receive more points per exchange than A and C. The D_N hypothesis also predicts B > (A,C). The null hypothesis predicts no difference in the point accumulations of B, A, and C.

3. The GPI indicates that D and E have equal power, and so should have a 12-12 division of points. D_N predicts that D will obtain higher profits than E. Our prediction can be falsified either by D > E, as D_N predicts, or by E > D.

4. E's profits will exceed those of A and C since E is in an equipower dyad and the others are low-power positions. The D_N and null hypotheses predict no profit differences among E, A, and C.

Results

In 100 negotiation rounds across five sessions, only three exchanges occurred between B and D.[13] The difference between this number and the null hypothesis of 16.667 was assessed with the z-test for proportions. The result, $z = 3.666$, $p < .0003$, supports Hypothesis 1 and refutes the null hypothesis.

Table 1 shows the average number of points per session for each position. B clearly obtained favorable exchange rates, above 19-5 in all but one session. The t-tests show

---

[13] The three B-D exchanges occurred in three different experimental groups, on second, third, and fourth rounds. In two cases, B received 12 points, in the third, 11. This indicates that the Bs were checking their alternatives, but quickly found no reason to continue such explorations.

that in every session, B's mean profits were significantly above 12 (and, by necessity, A's and C's significantly below). The null hypothesis is rejected and Hypothesis 2 and the D_N prediction are supported.

Table 1 shows that the mean D-E exchange rates for each session differed only slightly from the 12-12 split; t-tests indicate that none of these differences was statistically significant. Therefore, Hypothesis 3 is also confirmed and the D_N hypotheses rejected.

As for Hypothesis 4, the mean point total for position E was 12.12, A's was 4.81, and C's was 4.91. Combining session means for the latter two positions and testing against E's scores, $t = 7.522$, $p < .0005$. Hypothesis 4 is supported and the null and D_N hypotheses refuted.

In sum, this study provided strong support for the $p(1)$ measure as tested against its null hypotheses and the revised vulnerability measure. In the next section we present $p(e)$, a generalization for *multi-exchange* networks, that is, networks in which actors exchange more than once per round.

DOMAINS OF POWER AND MULTI-EXCHANGE NETWORKS

Identifying Domains

The concept of *domain* simplifies GPI calculations in multi-exchange networks. Domains are independent subnetworks—independent in the sense that structural changes in one cannot affect power in another.

First, let e be the maximum number of *unique exchanges* that positions can make in a given round. Two exchanges are unique for i only if they involve different relations. To identify domains we will need to distinguish e^+ and e^- positions: e^+ positions have more than e relations, and e^- positions have e or fewer. In Figures 3–5, e^+ positions are boxes, e^- positions are circles.

There are two types of domains. A *dyadic domain* is two related e^- positions. A *power domain* is a set of one or more related e^+ positions, along with all e^- positions related to any member of this set. Formally,

> DOMAINS: Given the set V of all positions on a path between i and j, i and j are in the same domain if and only if there exists a path such that either (1) $V = \{ \}$, or (2) all members of V are e^+ positions.

For example, both positions in the one-exchange network of Figure 3a are in the same domain since the set of positions (V) on the path connecting them is empty. They form a dyadic domain. Network 3b, in which $e = 1$, forms a single-power domain: all pairs of positions are either related or can be reached through a path containing only e^+ positions (boxes). Network 3c is also a single-power domain and, as noted earlier, no position has a structural advantage. This shows that being an e^+ position is necessary but not sufficient to produce high power (Willer and Patton 1987). Network 3d also forms a single-power domain.

By comparing 3c to 3d, we can see how a change in one part of the power domain can have distal effects. Note that 3d is the 3c network with E added to the D position. In 3c, A was in an equipower relation with B. But A becomes a low-power position when E is attached. In fact, the relative power of positions in all relations in the network change when E is added.

We can draw two implications at this point. (1) If there is differential power in a domain, then there is an e^+ position. This yields the useful contrapositive assertion: the absence of e^+ positions implies no power differentiation. So for power to exist in a domain (or in a network, for that matter), at least one position must have an excess of available partners. (2) All one-exchange networks form single domains. The reason will be clear as we next show that when $e > 1$, a network can have multiple domains.
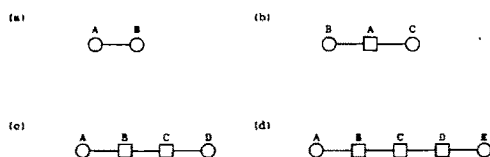
When $e > 1$, by the unique-exchange restriction, a position can exchange e times only if it has e or more relations. Some positions—those with fewer than e relations—can have effective maxima less than e. Since $e = 2$ in Figure 4a, for example, A can exchange twice, but B and C have effective maxima of one.

In Figure 4a, B-A-C now has two dyadic domains, (AB) and (AC); there is no core of one or more e^+ positions. By the assertion given above, since there are no e^+ positions, there is no power differentiation. This is reasonable since neither B nor C is excluded from exchanging with A in a given round. No position has excess exchange opportunities, and no position may garner favorable profit divisions. The same logic holds, in fact, for chains of any length, including the 4b network. This network contains four dyadic domains.

The manifestation of distal effects depends on the extent of domains. For instance, since B and C in 4a are in different domains, neither removing C nor adding new relations to C can affect B's power, and vice versa. The same is true for any two positions, e.g., B and D, lying in different domains in the 4b chain. In contrast, 4c shows that attaching F to the center of the 4b chain changes C from an e^- to an e^+ position—from a circle to a box. This creates a (BCDF) domain. C now has power over B, D, and F since it can exclude one of them in each round. Attaching a new position to D would remove C's power and benefit B, further demonstrating that B and D are in the same domain.

Calculating p(e)

Every position in a multi-exchange network will have a p index for each of its domains: $p_{id}(e_d)$ is position i's power in domain d, under the condition that i can make e_d exchanges per round within this domain.

Let m_{idk} be the number of nonintersecting paths of length k from position i in domain d,

Fig. 3. $e = 1$
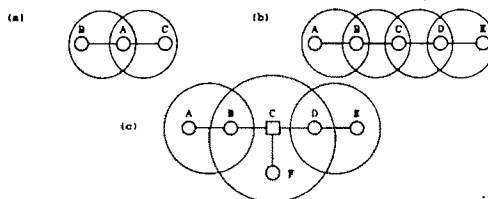
Fig. 4. $e = 2$

and h the longest such path from i in that domain. Only paths within a domain's boundaries are counted. As illustrated in the graphs, each path begins and ends with circles, between which there are either no positions or only boxes. Position i's GPI within the domain is[14]

$$p_{\text{id}}(e_{\text{d}}) = [1/e_{\text{d}}] \sum_{k=1}^{h} (-1)^{(k-1)} m_{\text{id}k} \quad (2)$$

$p(e)$ is closely related to $p(1)$ and similarly calculated. Multiplying the summation by $1/e_{\text{d}}$ simply places $p(e)$ and $p(1)$ values on the same scale.

Let us apply equation 2 (which now substitutes for Axiom 1) to network 4a, with $e = 2$. The two dyadic domains are indicated by (AB) and (AC) subscripts. We see that $p_{\text{B}} = p_{\text{C}} = p_{\text{A(AB)}} = p_{\text{A(AC)}} = (1/1)(1) = 1$. Each position has, in each of its domains, exactly one one-path and one exchange. Therefore, A has no power advantage in either of its domains. Similar results obtain in Figure 4b.

The 4c network has (AB) and (DE) dyadic domains and power domain (BCDF). Again, $p = 1$ for members of dyadic domains. However, for the power domain we calculate $p_{\text{C(BCDF)}} = (1/2)(3) = 3/2$, and for B, D, and F, $p = (1/1)(1-1) = 0$. Thus, C has a power advantage in both of its exchanges, B and D have low power in one of their exchanges and equal power in the other, A and E have equal power in their one exchange, and F has low power in its one exchange.

We may also calculate an *average power index*, $\bar{p}_i$, as the mean of i's indices across domains. In 4c, $\bar{p}_c = 3/2$; $\bar{p}_A = \bar{p}_E = 1$; $\bar{p}_B = \bar{p}_D = (1 + 0)/2 = .5$; $\bar{p}_F = 0$.

The Figure 5a network is the same as Figure 2, but redrawn using the circle and box notation. When $e = 1$, the network is a single domain and only the Es are high-power positions. In 5b, where $e = 2$, the situation is drastically altered. Only D has power advantages, with the Es all having low power relative to D. Furthermore, the E-F relations form three equipower dyadic domains.

Fig. 5

The 5a and 5b networks tested the GPI generalization. The two networks have identical shapes. Only the number of exchanges per round differs. Cook et al.'s (1983) simulations found the Es to be high-power positions in this network; $p(e)$ concurs, but *only* for the special case of $e = 1$.

EXPERIMENT 2

Experiment 2 tests the Figure 5 networks under $e = 1$ and $e = 2$ conditions. In spite of their identical shape, our analysis indicates that these networks should exhibit radically different profit distributions.

Method

Procedures for this experiment were similar to those used in Experiment 1. In this case, however, each subject negotiated from the different network positions under both one-exchange and two-exchange conditions, controlling for any personal characteristics of subjects that might confound the test.

Instructions for the one- and two-exchange conditions were identical, save for the

---

[14] For clarity, i subscripts have been suppressed for the e and h variables, d is suppressed for h, and $p_{\text{kd}}(e_{\text{d}})$ will be written as $p(e)$ or p. Note that Axiom 1 is now comprised of the more general equation (2).

number of exchanges allowed per round. In the two-exchange condition only, D and E could exchange with up to two different partners in the same round.

Four groups were run. Each group had seven subjects, one for each of the seven network positions. Two of the groups had the one-exchange condition first, followed by the two-exchange condition. The other two groups had the order of conditions reversed. As in the previous experiment, each subject occupied each network position over a series of four negotiation rounds. The design produced a total of 224 negotiations, 112 under each exchange condition.

After completing both parts of the experiment, subjects were paid according to the number of points they had accumulated, around $7.00 on average.

Hypotheses

The following hypotheses apply to the Figure 5 networks. All are tested against the null hypotheses that every relation would average 12-12 divisions.

1. In the one-exchange condition, the Es will exercise power over the Fs and D, and so the Es will all receive higher point totals than the others.

2. In the two-exchange condition, only D will exercise power. D will obtain higher point accumulations than the Es.

3. In the two-exchange condition, Es exchange in two domains. In the power domain, they will receive unfavorable profit divisions with D. In their respective dyadic domains, they will receive 12-12 divisions with the Fs.

Results

Table 2a and Figure 6a show the mean number of profit points obtained by each position under the one-exchange condition. The position labels for Figure 5a are also the
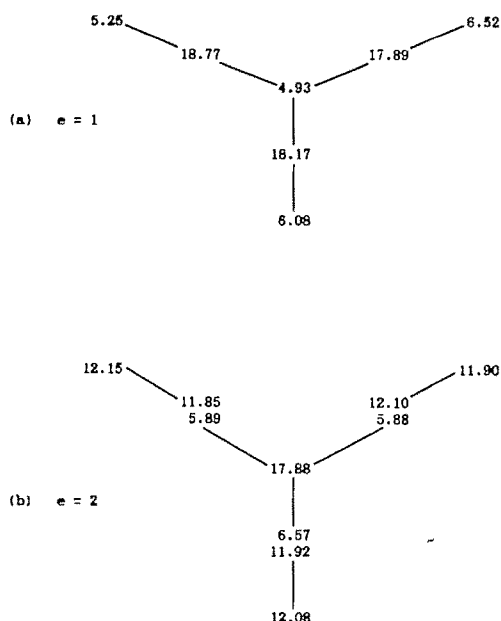
Fig. 6. Results of a Second Experiment

column headings of Table 2a. The Es clearly obtained favorable profits, around an 18-6 split on the average. The t-tests show that the Es' profits were significantly greater than a 12-12 split. Moreover, the Es exercised power over both D and the Fs. Hypothesis 1 is supported.

Table 2b and Figure 6b show results for the two-exchange condition. Now the power relationships have been reversed from the one-exchange condition, with the Es losing power and D gaining. As was the case for the Es under the one-exchange condition, D was able to gain approximately 18-6 profit divisions—significantly greater than the 12-12 split. Hypothesis 2 is confirmed.

Hypothesis 3 predicted equipower relations between the Es and their adjacent Fs under two-exchange conditions. The Es and the Fs should then have 12-12 profit-point divisions. As Table 2b shows, the 12-12 split was approximated. None of the differences were significant. Hypothesis 3 is also supported.

Table 2a. Second Experiment: Profit by Position, One-Exchange Condition

	Position						
	D	E_1	E_2	E_3	F_1	F_2	F_3
mean profit	4.93	18.77	17.89	18.17	5.25	6.52	6.08
s.d.	4.55	4.58	4.82	4.64	4.56	4.86	4.65
t	13.65	15.53	12.91	13.87	13.65	10.64	11.66

Note: All tests significant at $p<.001$, 1-tailed.

Table 2b. Second Experiment: Profit by Domain and Position, Two-Exchange Condition

Power Domain	D-E$_1$	D-E$_2$	D-E$_3$
mean profit	18.05	18.12	17.43
s.d.	4.61	4.50	4.65
t	11.66	11.63	9.77

Note: D's profit shown, E's profit = 24 − D's. All results significant at $p<.001$, 1-tailed.

Dyadic Domains	E$_1$-F$_1$	E$_2$-F$_2$	E$_3$-F$_3$
mean profit	12.15	11.90	12.08
s.d.	1.46	1.46	.73
t	1.05	.72	1.17

Note: E's profit shown, F's profit = 24 − E's. No significant test results.

In sum, this experiment provided strong support for the hypotheses testing the GPI generalization to multi-exchange networks. The presence of domains within the larger network under the two-exchange conditions strongly influenced the exchanges transpiring within those domains. As far as we know, such phenomena are not anticipated by alternative network-exchange theories.

NEW THEORETICAL DIRECTIONS

In addition to making its predictions more precise, the formality of our theory has made it easier to develop extensions. We briefly note five that are in varying stages of development and corroboration.

M-Exchange Networks

After developing the $p(e)$ model, we discovered that, with no loss of precision, different positions may seek different maximum numbers of exchanges per round—what we call the "M-Exchange" condition. This admits networks in which actors may seek exchange in one, some, or all of their relations. No reformulation of the GPI is needed to deal with this extension. The analysis predicts a new class of previously unanticipated power shifts.

No-Round Exchange

Allowing nonunique exchanges lets positions exchange more than once per round *within relations*. This effectively eliminates the need for exchange rounds. This is the *M-Exchange No-Round* condition. Now i may exchange up to e_i times or until those to which it is related

have exchanged up to their limits. Though this is a more complex situation, it is still true that only e^- positions can have low power and only e^+ positions can have high power.

Resource-Pool Values

If resource pools are different sizes in different relations, then there is another source for network breaks (Bonacich 1987b; Willer and Patton 1987). For example, in the one-exchange B-A-C network, let B and A negotiate over the division of 30 points, while A and C negotiate over 10. At first A will benefit from the bids of B and C. Eventually C will offer 9 points to A, keeping 1. Then B will offer 10 to A, keeping 20. C cannot meet this bid and still receive profit. Therefore, exchange should continue exclusively between B and A, with C excluded from the network. With the loss of C, only an A-B dyad remains and profits should reach a 15-15 split. Thus, power relations can be affected by variations in resource-pool values.

Flow-Networks

So far, we have focused on exchange conditions under which resources cannot transfer across relations. We have done so primarily because this is the condition under which most of the relevant research was conducted. However, as others have indicated (Marsden 1983; Bonacich 1987a), it is worthwhile to relax this restriction and consider networks with transferable resources—those in which resources may flow through positions.

A consequence of extending into the realm of flow-networks is that positions may have power over others to which they are not related, depending on the initial distributions of resources and on which actors seek which resources. This is similar to Marsden's view. The foremost difference between that view and our approach is that we incorporate explicit assumptions about individual negotiation strategies and the conditions of exchange—factors that affect exchange outcomes, breaks, and domains.

Positive Connection

While negative connections place an upper limit on the number of exchanges in which a

"hub" position may engage per round, positive connections place a lower limit on the number of exchanges in which the position must engage to realize a profit (Patton 1986). An example is the manufacturer who must obtain *all* components for a synthetic product before that product becomes a viable source of revenue. New research shows that the exchange dynamics that occur in positive connections differ markedly from those in negatively connected networks, and power advantages belong to peripheral positions in branches such as B-A-C (Patton and Willer 1987). This work on positive connections only begins to uncover a range of phenomena at least as broad and interesting as those associated with negative connections.

CONCLUSION

Our findings indicate that by only focusing on the effects of networks *per se,* alternative network theories do not recognize that power and resource distributions depend as much on prevailing exchange conditions as they do on configurations of positions and relations. We introduced a model that considers both structural form and exchange conditions, anticipating and explaining such phenomena as relative power, network breakage, power reversals and domain-specific effects. The studies that we described are only the first of many that could investigate stability and instability in exchange networks.

Future developments aside, we have found the present incarnation of the theory quite useful for understanding many real-world power struggles in exchange networks—from international disputes over geographical control to toddlers' negotiations over the sharing of playthings. Whatever the application, the theory directs us to specify the relevant actors and resources, identify other pertinent relations in which the actors are engaged, observe who seeks exchange with whom, identify which actors risk exclusion from valued resources, consider temporal constraints such as ultimatums or deadlines that create exchange rounds and, in general, determine the

extent to which the exemplar departs from the idealized scope conditions of the theory.

Our work also has implications for two very general questions that are relevant to structural approaches: (1) what is the appropriate unit of analysis for structural theories; and (2) how are characteristics of structures and the social units within them mutually determined?

Regarding the first question, we eschew the designation of one unit of analysis as, in general, more or less appropriate than another. Our theory explains certain actor and network behaviors. In any given instance, the network may be an organization, as may the actor. It follows that actors may or may not be individual persons. All that matters is that the units considered have the necessary properties. Therefore, no unit of analysis is generally most appropriate for structural approaches.

We can offer no universal solution to the question of how social structures and constituent units each determine properties of the other. Our approach does, however, point to *excludability* as a linchpin securing individual and network realms. That is, structures and exchange conditions at times bar some actors from procuring the resources they value and desire. Thus, power *happens* to those whose positions allow them to dodge the struggle to avoid exclusion.

As the foregoing review of extensions-in-progress implies, we do not claim that our theory is finished or unimprovable. Nor do we claim that it explains all phenomena within the purview of alternative formulations. It is, however, consistent with the findings of *all* previous experimental research on exchange networks. Moreover, it addresses a range of conditions and generates predictions that are either beyond the range of alternative formulations or simply contradict them, depending upon how one interprets their scope. Our long-term goal is to continue incremental extensions and systematic tests of increasingly refined network-exchange models.

APPENDIX

The Axiomatic Theory

<u>Symbols</u>:

i, j	:	actors in relation $i\text{---}j$
V	:	all positions related to i, other than j
Z	:	all positions related to j, other than i
v	:	a member of V
z	:	a member of Z
r_{ij}	:	resources received by i from j from an exchange
p_i	:	i's power index
p_{ij}	:	$p_i - p_j$
E_{ij}	:	i and j exchange
S_{ij}	:	i seeks exchange with j
k	:	empirical constant

<u>Logical Operators</u>:

x & y	:	conjunction ("x and y")
x or y	:	inclusive disjunction ("x and/or y")
-x	:	negation ("not x")
x ---> y	:	implication ("If x, then y")
x <--> y	:	biconditional ("x if and only if y")
(x)	:	universal quantifier ("For all x such that ...")
(∃x)	:	existential quantifier ("There is an x such that ...")

<u>Scope Conditions for Relations</u>

SC_5. $(i)(j)(\exists v)(S_{iv} \text{ or } S_{ij})$

SC_6. $(i)(j)[E_{ij} \longrightarrow (r_{ij} + r_{ji} = k)]$, $k > 0$;
         $(i)(j)[-E_{ij} \longrightarrow (r_{ij} + r_{ji} = 0)]$

SC_7. $(i)(j)\{E_{ij} \longleftrightarrow [(r_{ij} > 0) \& (r_{ji} > 0)]\}$

<u>Axioms</u>

A_1. $(i)(p_i = ...)$ (see equations 1 and 2 in text)

A_2. $(i)(j)\{S_{ij} \longleftrightarrow [(p_{ij} > p_{ji}) \text{ or } (v)(p_{ij} \geq p_{iv})]\}$

A_3. $(i)(j)[E_{ij} \longrightarrow (S_{ij} \& S_{ji})]$

A_4. $(i)(j)\{E_{ij} \longrightarrow [(p_i > p_j) \longleftrightarrow (r_{ij} > r_{ji})]\}$

<u>Theorems</u>

T_1. $(i)(j)[(V = \{ \}) \longrightarrow S_{ij}]$

T_2. $(i)(j)[(-S_{ij} \text{ or } -S_{ji}) \longrightarrow -E_{ij}]$

T_3. $(i)(j)\{-S_{ij} \longleftrightarrow [(p_{ij} \leq p_{ji}) \& (\exists v)(p_{ij} < p_{iv})]\}$

T_4. $(i)(j)\{[(p_{ij} \leq p_{ji}) \& (\exists v)(p_{ij} < p_{iv})] \text{ or }$
          $[(p_{ji} \leq p_{ij}) \& (\exists z)(p_{ji} < p_{iz})] \longrightarrow -E_{ij}\}$

<u>Proofs</u>*

<u>Theorem 1</u>

(1)	$V = \{ \}$	premise (P)
(2)	$-S_{iv}$	(1), definition of V
(3)	$-(\exists v)S_{iv}$	(2), Interchange of Quantifiers (IQ)
(4)	$(\exists v)(S_{iv} \text{ or } S_{ij})$	SC_5

| | S_{ij} | (3), (4), <u>disjunctive syllogism</u> |

* For clarity, most universal quantifiers have been suppressed.

Theorem 2

(1) $-S_{ij}$ or $-S_{ji}$ P
(2) $-(S_{ij} \& S_{ji})$ (1), DeMorgan's Law (DL)
(3) $E_{ij} \longrightarrow (S_{ij} \& S_{ji})$ A_3
_____
     $-E_{ij}$ (2), (3), modus tollens (MT)

Theorem 3

(1) $-S_{ij}$ P
(2) $[(p_{ij} > p_{ji})$ or $(v)(p_{ij} \geq p_{iv})] \longrightarrow S_{ij}$ A_2, Biconditional Law (BL)
(3) $-[(p_{ij} > p_{ji})$ or $(v)(p_{ij} \geq p_{iv})]$ (1), (2), MT
(4) $-(p_{ij} > p_{ji}) \& -(v)(p_{ij} \geq p_{iv})$ (3), DL
(5) $-(p_{ij} > p_{ji})$ (4), Law of Simplification
(6) $-(v)(p_{ij} \geq p_{iv})]$ (4), Law of Simplification
(7) $p_{ij} \leq p_{ji}$ (5), Law for Inequalities
(8) $(\exists v)(p_{ij} < p_{iv})$ (6), IQ
_____
     $(p_{ij} \leq p_{ji}) \& (\exists v)(p_{ij} < p_{iv})$ (7), (8), Law of Adjunction

Theorem 4

(1) $[(p_{ij} \leq p_{ji}) \& (\exists v)(p_{ij} < p_{iv})]$ or   P
     $[(p_{ji} \leq p_{ij}) \& (\exists z)(p_{ji} < p_{iz})]$
(2) $(p_{ij} \leq p_{ji}) \longleftrightarrow -(i)-(j)(p_{ij} > p_{ji})$ Law for Inequalities (LI),
                                                                           Double Negation (DN)
(3) $(p_{ji} \leq p_{ij}) \longleftrightarrow -(j)-(i)(p_{ji} > p_{ij})$ LI, DN
(4) $(\exists v)(p_{ij} < p_{iv}) \longleftrightarrow -(v)(p_{ij} \geq p_{iv})$ LI, DN, IQ
(5) $(\exists z)(p_{ji} < p_{jz}) \longleftrightarrow -(z)(p_{ji} \geq p_{jz})$ LI, DN, IQ
(6) $[-(p_{ij} > p_{ji}) \& -(p_{ij} \geq p_{iv})]$ or   (2), (3), (4), (5) substituted
     $[-(p_{ji} > p_{ij}) \& -(p_{ji} \geq p_{jz})]$ into (1)
(7) $[-(p_{ij} > p_{ji}) \& -(p_{ij} \geq p_{iv})] \longrightarrow$ DL
     $-[(p_{ij} > p_{ji})$ or $(p_{ij} \geq p_{iv})]$
(8) $[-(p_{ji} > p_{ij}) \& -(p_{ji} \geq p_{jz})] \longrightarrow$ DL
     $-[(p_{ji} > p_{ij})$ or $(p_{ji} \geq p_{jz})]$
(9) $-[(p_{ij} > p_{ji})$ or $(p_{ij} \geq p_{iv})]$ or   (6), (7), (8)
     $-[(p_{ji} > p_{ij})$ or $(p_{ji} \geq p_{jz})]$
(10) $S_{ij} \longleftrightarrow [(p_{ij} > p_{ji})$ or $(p_{ij} \geq p_{iv})]$ A_2
(11) $S_{ji} \longleftrightarrow [(p_{ji} > p_{ij})$ or $(p_{ji} \geq p_{jz})]$ A_2
(12) $(S_{ij} \& S_{ji}) \longleftrightarrow$ (10), (11)
        $\{[(p_{ij} > p_{ji})$ or $(v)(p_{ij} \geq p_{iv})] \&$
        $[(p_{ji} > p_{ij})$ or $(z)(p_{ji} \geq p_{jz})]\}$
(13) $-(S_{ij} \& S_{ji})$ (9), (12)
(14) $-S_{ij}$ or $-S_{ji}$ (13), DL
_____
     $-E_{ij}$ (14), T_2

REFERENCES

Blalock, Hubert M. and Paul H. Wilken. 1979. *Intergroup Processes*. New York: Free Press.

Blau, Peter. 1964. *Exchange and Power in Social Life*. New York:Wiley.

Bredemeier, Harry C. 1978. "Exchange Theory." In *A History of Sociological Analysis*, edited by T. Bottomore and R. Nisbet. New York: Basic Books.

Bonacich, Phillip. 1987a. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92:1170–82.

_____. 1987b. Personal communication.

Burgess, R.L. and T.L. Huston, eds. 1979. *Social Exchange in Developing Relationships*. New York: Academic Press.

Cook, Karen S., ed. 1987. *Social Exchange Theory*. Newbury Park, CA: Sage.

Cook, Karen S. and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43:721–39.

Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89:275–305.

Cook, Karen S., Mary R. Gillmore, and Toshio Yamagishi. 1986. "Point and Line Vulnerability as Bases for Predicting the Distribution of Power in Exchange Networks: Reply to Willer." *American Journal of Sociology* 92:445–48.

Earle, Timothy, and Jonathon Ericson. 1977. *Exchange Systems in Prehistory*. New York: Academic Press.

Ekeh, Peter P. 1974. *Social Exchange Theory: The Two Traditions*. Cambridge, MA: Harvard University Press.

Elkin, A.P. 1953. "Delayed Exchange in Wabag Sub-District, Central Highlands of New Guinea." *Oceania* 23:161–201.

Emerson, Richard M. 1972a. "Exchange Theory, Part I: A Psychological Basis for Social Exchange." In *Sociological Theories in Progress*, vol. 2, edited by J. Berger, M. Zelditch, Jr., and B. Anderson. New York: Houghton Mifflin.

_____. 1972b. "Exchange Theory, Part II: Exchange Relations and Network Structures." In *Sociological Theories in Progress*, vol. 2, edited by J. Berger, M. Zelditch, Jr., and B. Anderson. New York: Houghton Mifflin.

_____. 1976. "Social Exchange Theory." *Annual Review of Sociology* 2:335–62.

_____. 1981. "Social Exchange Theory." In *Social Psychology: Sociological Perspectives*, edited by M. Rosenberg and R. Turner. New York: Basic Books.

Fararo, Thomas J. 1973. *Mathematical Sociology*. Huntington, NY: Kreiger.

Gergen, Kenneth. 1969. *The Psychology of Behavior Exchange*. Reading, MA: Addison-Wesley.

Galaskiewicz, J. 1985. "Interorganizational Relations." *Annual Review of Sociology* 11:281–304.

Gilham, Steven. 1981. "State, Law, and Modern Economic Exchange." In *Networks, Exchange and Coercion*, edited by D. Willer and B. Anderson. New York: Elsevier/Greenwood.

Hansen, Knud L. 1981. " 'Black' Exchange and its System of Control." In *Networks, Exchange and Coercion*, edited by D. Willer and B. Anderson. New York: Elsevier/Greenwood.

Harary, Frank. 1969. *Graph Theory*. Reading, MA: Addison-Wesley.

Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. 1965. *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.

Heckathorn, Douglas. 1980. "A Unified Model for Bargaining and Conflict." *Behavioral Science* 25:261–84.

Heath, Anthony. 1976. *Rational Choice and Social Exchange: A Critique of Exchange Theory*. Cambridge: Cambridge University Press.

Homans, George C. 1974. *Social Behavior: Its Elementary Forms*. New York: Harcourt Brace Jovanovich.

Kahan, James P. and Amnon Rapoport. 1984. *Theories of Coalition Formation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhn, Alfred. 1974. *The Logic of Social Systems*. San Francisco: Jossey-Bass.

Lind, Joan. 1987. "Exchange Processes in History." *Sociological Quarterly* 28:223–46.

Loukinen, Michael. 1981. "Social Exchange Networks." In *Networks, Exchange and Coercion*, edited by D. Willer and B. Anderson. New York: Elsevier/Greenwood.

Markovsky, Barry. 1987a. "Toward Multilevel Sociological Theories: Simulations of Actor and Network Effects." *Sociological Theory* 5:100–15.

_____. 1987b. "NETSIM." Program available from the author.

Markovsky, Barry, David Willer, and Travis Patton. 1987. "Power in Complex Exchange Networks." Available from authors.

Marsden, Peter V. 1983. "Restricted Access in Networks and Models of Power." *American Journal of Sociology* 88:686–717.

_____. 1987. "Elements of Interactor Dependence." In *Social Exchange Theory*, edited by K.S. Cook. Newbury Park, CA: Sage.

Patton, Travis. 1986. "A Theoretical Analysis of Positively Connected Exchange Networks." Presented at the Annual Meetings of the Midwest Sociological Society, Des Moines, Iowa.

Patton, Travis and David Willer. 1987. "Power in Centralized Networks: The Conditions of Positive, Negative and Null Connection." Paper presented at the Annual Meetings of the American Sociological Association, Chicago.

Peplau, Letitia Ann. 1979. "Power in Dating Relationships." In *Women: A Feminist Perspective*, 2nd ed. edited by J. Freeman. Palo Alto, CA: Mayfield.

Polanyi, Karl. 1944. *The Great Transformation*. Boston: Beacon Press.

Sahlins, Marshall. 1972. *Stone Age Economics*. New York: Aldine.

Shubik, Martin. 1982. *Game Theory in the Social Sciences*. Cambridge, MA: MIT Press.

Skinner, Steven J. and Joseph P. Guiltinan. 1986. "Extra-network Linkages, Dependence, and Power." *Social Forces* 64:702–13.

Stolte, John and Richard Emerson. 1977. "Structural Inequality: Position and Power in Network Structures." In *Behavioral Theory in Sociology*, edited by R. Hamblin and J. Kunkel. New Brunswick, NJ: Transaction Books.

Thibaut, John and Harold H. Kelley. 1959. *The Social Psychology of Groups*. New York: Wiley.

Turner, Jonathan H. 1986. *The Structure of Sociological Theory*, 4th ed. Chicago: Dorsey.

Wagner, David G. and Joseph Berger. 1985. Do Sociological Theories Grow? *American Journal of Sociology* 90:697–728.

Walker, Henry A. and Bernard P. Cohen. 1985. "Scope Statements: Imperatives for Evaluating Theory." *American Sociological Review* 50:288–301.

Willer, David. 1981. "Quantity and Network Structure." In *Networks, Exchange and Coercion*, edited by D. Willer and B. Anderson. New York: Elsevier/Greenwood.

———. 1986. "Vulnerability and the Location of Power Positions." *American Journal of Sociology* 92:441–48.

———. 1987. *Theory and the Experimental Investigation of Social Structures*. New York: Gordon and Breach.

Willer, David and Bo Anderson, eds. 1981. *Networks, Exchange and Coercion*. New York: Elsevier/Greenwood.

Willer, David and Barry Markovsky. 1986. "Exchange and Power in Social Networks: Theory and Proposed Research." Sociology Working Paper Series, U. of Iowa.

Willer, David, Barry Markovsky, and Travis Patton. Forthcoming. "The Experimental Investigation of Social Structures." In *Sociological Theories in Progress*, vol. 3, edited by J. Berger and M. Zelditch, Jr.

Willer, David and Travis Patton. 1987. "The Development of Network Exchange Theory." In *Advances in Group Processes*, vol. 4, edited by E. J. Lawler and B. Markovsky. Greenwich, CT.:JAI Press.

WHEN BIGGER IS BETTER: DIFFERENCES IN THE INDIVIDUAL-LEVEL EFFECT OF FIRM AND ESTABLISHMENT SIZE*

WAYNE J. VILLEMEZ
Louisiana State University

WILLIAM P. BRIDGES
University of Illinois-Chicago

Building on previous demonstrations of a linkage between organizational size and individual outcomes, this paper employs a matched employer-employee data set to investigate that relationship. Some portion of the size effect on worker earnings is found to be indirect, as many have assumed, derived either through the organizational consequences of size or the industrial/market consequences of scale. However, the effect is not straightforward. It varies by group (gender, occupation, and industry), is differentially mediated by internal labor markets, literacy requirements, and unionization, differs according to whether firms are single- or multi-establishment, and for individuals as to whether firm size or establishment size is most salient to their earnings. For white-collar workers, firm size is most important, but not primarily due to organizational aspects of size. For blue-collar workers, establishment size is most important, largely due to the organizational concomitants of size. Female clerical workers are found to be a special case, and clear gender differences exist in the effect of size.

INTRODUCTION

Increasingly, researchers in social stratification have recognized that properties of organizations may play a crucial role in explaining the nature of social inequality. Stolzenberg's (1978) influential paper showed how the size of the organization affects individual workers' labor-market outcomes. Despite studies that have replicated various size-of-employer effects (e.g., Kalleberg, Wallace, and Althauser 1981; Baron and Bielby 1984; Hodson 1984), no one has satisfactorily explained this pattern. Although there have been some noteworthy beginnings (Baron and Bielby 1980; Granovetter 1984), the data required to investigate this problem—information on individual employees, their individual employers, and their industries—have been scarce, hampering progress. This is troubling because the size-of-employer effect lies precisely on the interstices between stratification and organizational theory, and its elaboration is essential to bridging that

crevice. In this paper, we analyze a data set that permits a more detailed delineation of the process by which employer size affects individual labor-market outcomes.

Beyond the fact of a size-of-employer effect on the labor-market outcomes of individuals, little else has been empirically verified. Some rudimentary questions have yet to be answered: (1) Is the size effect simply a crude and imprecise proxy for more important determinants of firm-level attributes? (Baron and Bielby 1980); (2) If so, which attributes of large organizations matter most for individual outcomes? (3) Is it partially a proxy for industrial sector, e.g., core/periphery? (Stolzenberg 1978); (4) Is it a uniform effect, or does it vary across categories of individuals, occupations and/or industries? (5) Is firm size crucial (Stolzenberg 1978; Kalleberg et al. 1981), or is establishment size more important (Granovetter 1984)? Some basic questions must be answered before we can proceed to broader issues. Five basic issues are the focus of this article.

Previous Research

The effect of size on aggregate worker outcomes is well known and clear-cut. Debate exists over the impact of size on subjective factors such as job satisfaction (see Hall 1982, 58–60 for a summary), but not on its impact on objective factors. Studies have

demonstrated, for example, that industries characterized by larger firms pay higher wages, as do larger firms within industries (see Masters 1969; Lester 1967; Phelps Brown 1977; Bailey 1980; Podgursky 1986; see Baron 1984, for an excellent review of much of this literature).

Several plausible explanations for the size effect on individuals were originally offered by Stolzenberg. Drawing on Child's (1973) demonstration that organizational size is substantially related to dimensions of organizational structure, he focused on the strong likelihood that two organizational dimensions, standardization and documentation, would influence individual attainment. He argued that higher levels of documentation would place a premium on the ability to communicate in writing, resulting in a higher return to education. Similarly, standardization of hiring and remuneration decisions would lead to a reliance on standardized criteria for decision making, which would also result in a higher return to schooling. A second argument adopted by Stolzenberg draws on the segmented-economy literature. He argued that since core firms are large and offer good, stable jobs and internal labor markets, and periphery firms are small and offer unstable low-wage jobs with little potential for advancement, achievement processes should differ. Stolzenberg described no explicit linkage between core/periphery firm differences and individual return to education. Perhaps he reasoned that better credentials are required to enter large firms through entry ports to internal labor markets, or, more simply, that core firms tend to be large firms that pay education better. Whatever the reason, he found a relationship between employer size and both worker-income return to education and occupational-status return to education.

These findings have been partially supported by Kalleberg et al. (1981), who, using the same data analyzed by Stolzenberg, showed that organization size has a significant positive impact on (log) earnings, net of industrial and occupational variables. They note, however, that education interacts negatively with size for females but not for males, which does not support Stolzenberg's findings. When the same equations were run for only white males and females, a positive size/education interaction appeared for white males, but the negative interaction remained for white females. Kalleberg et al. concluded

that Stolzenberg's findings of superior return to education in large organizations applied only to white males and that females have a superior return to education in smaller establishments.

Hodson's (1984) findings lend some support to Kalleberg et al., in that he finds a substantial size effect on earnings and differential effects for males and females. Hodson suggests that there is some support to conclude ". . . that men derive economic advantage primarily from the industrial environment associated with capital-intensive production systems whereas women derive advantage primarily from the bureaucratic rules associated with large company size" (1984, p. 345). Granovetter (1984) summarized some previous "size effect" literature, and argued that ". . . to the extent one's interest . . . lies in . . . what kind of labor-market situation workers find themselves in . . . establishments rather than firms are the appropriate unit" (1984, p. 328). He contended that below the middle-management level there is little inter-establishment movement within firms and so the single establishment ". . . typically defines the internal labor market in which their career line can progress" (p. 328). This argument ignores the fact that regardless of establishment size it is possible that firm size has a very strong effect on "kind of labor market situation" for individuals. Previous researchers have not suggested that individual movement among establishments gives firm size its effect, but that organizational constraints and scale effects do. But Granovetter's argument, slightly recast, could be useful for specifying expected differences in size effects for different types of workers. His argument implies that establishment size can have an effect *independent* of firm size. Also implicit in his argument is a point that runs counter to his major thrust (that establishment size is more important than firm size): firm size and establishment size could have different effects on different classes of workers.

Granovetter makes another important point. He suggests that the size/individual outcome relationship might be strongest for those in manufacturing and weakest (or nonexistent) for those in the expanding service industries. Contending that the arguments to explain the peripheral location of small manufacturing concerns do not readily generalize to other industries, he points out:

. . . it is unclear to what extent establishment size can be associated with a core-periphery distinction in trade or services. In this connection, . . . while average weekly wages of employees in manufacturing rise almost monotonically with establishment size—as suggested by dual-economy arguments—there is almost no correlation between these wages and establishment size in services. . . . These relations are zero order, and size could still have the expected positive correlation with wages net of such factors as occupation or education. (1984, p. 331)

Although these previous studies offer some conflicting conclusions, there are also areas of agreement. First, organizational size has some individual-level effects that probably vary across different population groups. Second, internal labor markets (ILMs) are an important covariate of organizational size, and internal labor-market participation helps to mediate the size-income relationship. In this regard, these studies are consistent with other literature that reports an association between organizational size and the presence of ILMs (Baron, Davis-Blake, and Bielby 1986; Oi 1983).

Given previous research, it may be that the superior earnings and earnings return to education in larger organizations could be due to one or several of the following attributes of large organizations: (1) the pervasiveness of internal labor markets; (2) the likelihood of being in the core sector of the economy; (3) the likelihood of being unionized; (4) the greater opportunity for being in a supervisory position; (5) prevailing formalized procedures that reward formal credentials; (6) more formal requirements for hiring, leading to greater reliance on formal credentials; (7) increased standardization and record-keeping, leading to higher demand for reading comprehension and writing skills (and thus formal education).

Hypotheses

We shall examine the distribution of each of these attributes to assess their appropriateness in a model that explains organizational-size effects on individual outcomes. Using appropriate intervening variables, it will be possible to test a number of hypotheses that follow from the literature.

Since organizational size clearly has some individual-level effects, we first hypothesize

that organizational size will have a positive impact on earnings net of education and experience. Based on an elaboration of Granovetter's argument that workers below the middle-management level often live out their careers in single establishments, we expect that establishment size would be most salient for blue-collar workers. For higher-level workers, there is both the possibility of inter-establishment movement and more lines of advancement in larger firms. Higher-level workers in large firms probably also have more contact with and are more visible to those in other firms than are similar workers in small firms. These advantages should also accrue to some extent to most other white-collar workers. Given this, we would expect firm size rather than establishment size to have more impact on the earnings of white-collar workers. Therefore, we hypothesize that firm size will have a positive impact on the earnings of white-collar workers, while establishment size will have a positive impact on the earnings of blue-collar workers.

One possible exception to the notion that only higher-level workers benefit from being in larger firms is the category of female clerical workers. If true, this is an important exception because in our sample clerical workers constitute over 35 percent of employed females (a proportion representative of the labor-force proportion). Kanter (1977, p. 74) has noted that "Secretaries derived their formal rank . . . from the formal rank of their bosses. . . . Most often, above the early grades, secretaries were not actually promoted at all on their own; they just remained with a boss who himself received a promotion." Since secretarial ladders are short (Kanter 1977, p. 71), we would expect secretaries to be better off where their bosses are likely to be better off. We would expect this even without the "mobility by attachment" mechanism suggested by Kanter's findings. Since the earnings of higher administrators are most often closely tied to the size of the enterprise they are running, internal equity and social comparison mechanisms should scale secretaries' wages to the salaries of those for whom they work. Thus, we hypothesize that firm size will have a positive impact on the earnings of female clerical workers.

If there are different types of size effects by occupation and different reasons for those effects, then the nature of the size-earnings

relationship should differ. For white-collar workers, if the size effect is primarily one of opportunity structure, the effect of size should remain strong net of controls for organizational and sectoral characteristics.[1] For blue-collar workers, however, size should be a proxy for establishment-level characteristics salient to earnings. We would expect the effect of organizational size on blue-collar earnings to be largely an indirect effect, operating through such characteristics. Formally hypothesized, individually salient organizational characteristics (ISOC) mediate the, effect of. organizational size on earnings for some, but not all, groups. For blue-collar workers, the effect of organizational size on earnings is primarily indirect, operating through ISOC. For white-collar workers, controlling for ISOC does not remove the positive effect of size on earnings, since the effect is primarily one of opportunity structure, and such structure is only partially determined through ISOC. Using the same reasoning, it also follows that controlling for ISOC will have the same effect for female clerical workers as for white-collar males.

The literature reviewed by Granovetter suggests that an organizational-size effect on individual earnings may occur only in manufacturing industries. · If true, this would mandate a prediction of no size effect in nonmanufacturing industries. However, this should not be true for white-collar workers, since Granovetter's case for the preeminent importance of establishment size seems applicable only to blue-collar workers. Further, it is not clear why firm size in any industry should not provide a broader opportunity structure for white-collar workers. Thus, we hypothesize that for blue-collar workers, organizational size (establishment size) will affect earnings only in manufacturing industries; this will not be true for white-collar workers, for whom firm size effects will occur across industries.

Recent work suggests that the effects of ILMs on earnings may not be as direct or as simple as the dual-economy literature portrays them. We argue that the "internality" of a

labor market is not always most fruitfully seen as a discrete variable and that degrees of internality are possible. We see as ideal types more formal definitions of internal labor markets that call for formal job ladders with entry only at the bottom and movement up according to clearly defined guidelines (see Althauser and Kalleberg 1981): ideal types realized in actual practice in only a minority of instances. Adopting such restrictive definitions creates two problems.

First, as a typological definition, this perspective defines the presence of ILMs by the joint occurrence of two related aspects: protection from outside competition and assurances of mobility by the delineation of a predetermined career path. However, these dimensions are separable, and in many industrial union settings protection from outside hiring is offered in conjunction with unplanned, but institutionalized, movement within the internal job system. This frequently takes the form of a "bidding system," in which movement is not the result of a prior plan but the consequence of individual preferences at a given point in time (see Burawoy 1979). The mobility aspects of ILMs can be treated separately from the protection aspect.

Second, jobs have degrees of protection, of potential for upward movement, and of formality in job ladders. That is, an informal hierarchy of jobs in which hiring into the middle and upper rungs was *usually* from the rung below would not meet a strict definition of an ILM, but may still give considerable advantage to those on that informal ladder. For another example, certain jobs may be recognized as de facto stepping-stones to higher level jobs in entirely different divisions or even establishments. These jobs may not be accessible only from the bottom rung of a ladder and may not be completely protected, but they would nevertheless provide an advantage to individual incumbents.

Treating ILMs as multidimensional variables allows for a clearer specification of their potential impact on earnings within firms. For example, we could predict that, with regard to the mobility dimension of ILMs, the more formal and certain the mobility chances, the lower the earnings associated with the job. This prediction follows straightforwardly from neoclassical economic theory: where training is firm-specific, employers and employees will share in the cost of undertaking this

---

[1] We do not argue against any effect of organizational structure on white-collar workers; "opportunity structure" might itself be captured at least partially by a variable such as "organizational complexity" (which we have not measured directly). Simply, organizational variables thought to be salient for all workers may not hold for white-collar workers.

training (Becker 1985; Cole 1979, pp. 38–40; Addison and Siebert 1979), but the apportionment of training costs between the parties is variable. The more risk the employee assumes in acquiring the training, the larger the portion of the costs borne by the employer (the risk derives from the employee's foregoing the opportunity to obtain training that has some value in the external market). With more structured mobility chances, the employee's risk is reduced and some training costs can be shifted to the employee in the form of a lower wage. Thus, for both the employer and the employee, there could be a tradeoff between the mobility component of an ILM and the employee wage level. Although this logic applies most clearly when training is firm specific, to the extent that the employer shares the cost of general training, a similar principle should apply, and the effect of mobility assurances on earnings will be negative.

The protection component of ILMs, on the other hand, should always be associated with higher earnings net of other factors. Controlling for the mobility aspect, protection from outside competition places the employee in the role of a partial monopolist vis-à-vis the employer and increases the employer's dependence on the existing work force. This should apply net of unionization and may in fact explain some portion of the association between union coverages and higher wages. In all cases we expect the effect of job protection on earnings to be positive.

We have presented four general hypotheses and now restate them formally.

1. Organizational size will have a positive impact on earnings net of education and experience.
    1a) Firm size will have a positive impact on the earnings of white-collar workers.
    1b) Establishment size will have a positive impact on the earnings of blue-collar workers.
    1c) Firm size will have a positive impact on the earnings of female clerical workers.
2. Individually salient organizational characteristics (ISOC) will mediate the effect of organizational size on earnings for some, but not all, groups.
    2a) For blue-collar workers, the effect of organizational size on earnings

will be primarily indirect, operating through ISOC.
    2b) For white-collar workers, controlling for ISOC will not remove the positive effect of size on earnings, since the effect is primarily one of opportunity structure, and such structure is only partially determined through ISOC.
3. For blue-collar workers, organizational size will affect earnings only in manufacturing industries; this will not be true for white-collar workers.
4a) For all workers, the effect of job protection on earnings will be positive.
4b) For all workers, the effect of mobility assurances on earnings will be negative.

DATA AND VARIABLES

The data in this paper are from a dual survey of the labor market in the Chicago SMSA, conducted in 1981. The surveyed were 2,712 employed persons (defined as those employed by others at least 20 hours per week) and a matched sample of their employers. Each interviewed worker was asked for his or her exact place of employment (name and street address). A second survey was then conducted, consisting of interviews with one or more high-level officials in each employing organization. The focuses of the second survey were the specific job of the original respondent, hiring procedures, and relevant organizational attributes. The second survey was supplemented by employer data purchased from Dun and Bradstreet and by industrial data gathered from published sources. For a detailed description of the sample and comparative data documenting its generalizability and representativeness, see Bridges and Villemez (1986a).

The main dependent variable, earnings, was derived from a question about earnings "from your main job before taxes and other deductions." For those few refusing to respond, a follow-up question allowed them to be placed in an earnings category (e.g., "more than $15,000 but less than $20,000), and we used the midpoint of that category.

We have separate measures of our major independent variables, firm and establishment size. Employers were asked about both sizes, employees were asked establishment size

only, and Dun and Bradstreet provided the same data (though much was missing). The employer-provided figures were assumed to be most up-to-date and reliable; when the employer failed to provide establishment size, it was estimated from the coefficients of an equation with the respondent's and Dun and Bradstreet's figures on the right-hand side: ESTAB = f(ESTAB[R] + ESTAB[D&B]). Given the distribution and previous practice, we use the natural log of size in all equations.

Our measure of internal labor markets is complex, based on the summation of three other compound variables. We define an ILM in terms of the extent to which a job is "protected" from outside competition, the extent to which a job is a "stepping-stone" to higher-level jobs, and the extent to which a job is part of a regularly defined ladder of jobs (informal "ladders" exist, and the second variable will capture those; the last captures formal ladders). We create three variables: protection, up-mobility, and ladder. For the variable protection, we define three levels: "protected," "sheltered," (protected, but not completely) and "neither protected nor sheltered." We have two reports on the protectedness of the job, one from the worker and one from the employer. Workers were asked, "Are most people who do the same job as you hired from within the organization, or are they hired from outside the organization?" The "within" reply was coded as protected, "about half and half" and similar responses were coded as sheltered, and "outside" was coded as neither. Employers were asked, "Are vacancies for the job of (read type of job) *usually* filled from outside (read name of company) or from within (read name of company)?" "Within" was coded as protected, "both" was coded as sheltered, and "outside" was coded as neither. In the most common case, where employee and employer agree, the coding is straightforward: protected = 2, sheltered = 1, neither = 0.[2]

For the up-mobility variable, we also use items from both employer and employee. The employee was asked, "Would you say that the job you have now is a stepping-stone to another job?" and, if yes, "Would it be a move within the same organization or to another organization?" Employers were asked, "Of those people who would still be with (read name of company) in ten years, what are the chances they will have been promoted to a higher-level job? Would you say no chance, 10 percent chance, 20 percent chance, or what?" If the employee said it was a stepping-stone job to another job within the organization *or* the employer said the probability of promotion was 80 percent+, up-mobile was coded 2. If the employee said it was a stepping-stone job, but to another job not necessarily in the same organization, *or* if the employer said the probability of promotion was only 51 percent–79 percent, up-mobile was coded 1. If the employer said the probability of promotion was 50 percent or less and the employee said it was not a stepping-stone job, up-mobile was coded 0.[3] Ladder was defined by one employer question: "Is this job of (refer to type of job) part of a regular series of positions by which people move to a higher level?" A yes was coded 1, a no 0.

Thus defined, the three were summed, resulting in the variable ILM, which ranges from 0 to 5 (protection 0–2 + up-mobile 0–2 + ladder 0–1). Each component will also be used separately.

For the variable formalization, we scored 1 to yes answers to each of the following, 0 otherwise, and summed: "For (read type of job) does your establishment . . . a) Conduct periodic wage and salary reviews?; b) Have supervisors prepare written performance evaluations that are kept on file?; c) Give company handbooks or manuals?; d) Prepare

Employer says	Employee says	Score
Protected	Protected	2
	Sheltered	2 if internal hire, else 1
	Neither	2 if internal hire, else 0
Sheltered	Protected	2 if internal hire, else 1
	Sheltered	1
	Neither	1 if internal hire, else 0
Neither	Protected	2 if internal hire, else 0
	Sheltered	1 if internal hire, else 0
	Neither	0

[2] Since not every case fell on the diagonal, to help resolve disagreements another variable was used. Workers were asked: "When you first came to work for your present employer, was it in the same job you have now or was it in a different job?" "A different job" means that the current one was an internal hire, evidence of protection. The coding follows:

[3] Our results are not substantively altered if we use any one of the components of either protection or up-mobility rather than the composite variable for each. This fact was also clearly evident in our LISREL findings (available upon request). We use the composite variables because we feel they are better measures than any single component and for clarity of presentation.

written job descriptions that define the duties of this type of job? e) Give job instructions in writing to employees on this type of job?" The variable ranges from 0 to 5.

The writing and reading variables were derived from two questions each: "Does your job require you to compose (to read) a lot of written material such as instructions, reports, memos, etc?" and "Is this an important part of your job?" (asked about each). The coding was as follows:

Writing (Reading) required	Important	Writing = (Reading =)
Yes	Yes	5
Sometimes	Yes	4
Yes	No	3
Sometimes	No	2
No	No	1

For the education and experience requirement variables, the following portions of a question to the employer were used: "To be eligible for (read type of job) is it required that the applicant . . . b) Have previous job experience in this line of work . . . e) Have a certain level of education or a particular degree?. . . e2) Would the applicant have had to have a certain level of education or a particular degree for a previous position to be eligible for this job?" A "yes" to b) gives a 1 to the experience required dummy variable; a "yes" to either e) or e2) gives a 1 to the education required dummy variable.

For the core/periphery variable, we use a dummy variable based on the industrial coding recommended by Tolbert, Horan, and Beck (1980), and score core as 1. For experience we use actual years of working experience. Other variables are either obvious or will be explained as they make their brief appearance below.

To avoid confounding effects at this early stage of investigation, we analyze data only for private sector employees (since "organizational size" is of dubious meaning in the public sector). There are an ample number of cases remaining after this restriction is imposed: for private-sector males, $N = 1,384$ (white collar, 771; blue collar, 613); for private-sector females, $N = 941$ (white collar, 699; blue collar, 242; clerical, 341; non-clerical, 600).

We have constructed a number of measures of key concepts and the validity of these

constructs has an important impact on the validity of our interpretations. As a check on our major constructed variables, we estimated LISREL models using the disaggregated multiple indicators of our key concepts. The results, available upon request, exactly duplicate those in the tables to follow.

ANALYSIS AND RESULTS

We begin with a partial replication of Stolzenberg (1978) and Kalleberg et al. (1981). Following Stolzenberg, for white males only we separately regress occupational status and earnings on education, experience, education squared, education × experience, and tenure for several categories of size (adding occupational status as an independent variable to the earnings equation). We then calculate the partial derivative of earnings with respect to education for each quartile of firm size. The following are the returns to education, from smallest to largest quartile: for earnings, -1196, 958, 850, 1353; for occupational status, 5.4, 5.3, 5.8, 4.8. Though we replicate his finding of a general upward trend for earnings, it is clearly not monotonic (the same pattern emerges when log earnings is the dependent variable). For occupational status, there is no discernible trend.[4]

Kalleberg et al. argued that if there was a superior income return to education in large firms, it should show up in an interaction between education and size. Regressing log(earnings) and then earnings on education, experience, log(firm size) and log(firm size) × education, we find the following coefficients of the interaction term: for log(earnings), white males .005 ($p < .02$), white females .003 (NS); for earnings, white males 255 ($p < .001$), white females 45 (NS). Though we do not find a negative effect for females, as did Kalleberg et al. (1981), we do replicate their basic conclusion: a positive

[4] Stolzenberg was forced to use establishment size as a proxy for firm size and to use precategorized size as well. Using establishment size by quartiles we find: for earnings, 606, 592, 1,628, 1,062; for occupational status, 5.1, 5.6, 4.9, 5.9. If we also use the five QES size categories we find: for earnings, 950, 1,248, −1,838, 1,419, 1,374; for occupational status, 6.7, 5.7, 6.0, 5.7, 6.2. In either case, our conclusions remain the same.

Table 1a. Means of Job-Relevant Variables by Quartiles of Firm Size, for Males, Private Sector.

	Firm Size			
	Small	2	3	Large
Earnings	21,726	21,705	25,869	28,274
Education	13.1	13.5	14.0	13.8
Experience	15.4	16.6	18.7	19.3
Internal	1.33	1.96	2.72	2.90
Protection	.44	.67	1.01	1.24
Up-mobility	.57	.81	1.00	1.06
Ladder	.31	.50	.67	.65
Reading	2.7	3.1	3.7	3.3
Writing	2.4	3.0	3.4	3.0
Union	.26	.29	.30	.44
Core	.59	.67	.70	.74
Authority	.91	1.00	.94	.84
Supervise others	.56	.59	.59	.53
Say in hiring/firing	.35	.38	.35	.32
Formalization	2.1	3.4	3.6	3.9
Education required	.31	.49	.49	.37
Experience required	.61	.63	.52	.50
Black	.08	.08	.10	.11

education-size interaction effect on earnings for males, and no such effect for females.

These preliminary findings suggest both a size effect on earnings and one that differs for males and females. We now examine the components of this effect, beginning with mean differences in characteristics across quartiles of firm size. We divide our sample by quartiles of firm size and within each quartile examine means of relevant organizational variables, as well as means of human capital and outcome variables.

As can be seen in Table 1, outcome variables as well as characteristics of the workplace vary systematically across catego-

ries of firm size. Mean earnings for males, for example, increases from $21,726 in firms of 1–106 persons to $28,274 in firms of 31,705+ employees. For females, the same increment is from $10,164 to $14,124, a larger increase, proportionately. Comparing Tables 1a and 1b, average education and experience increases with firm size for males, but not for females. For both, the likelihood of an internal labor market increases across quartiles. The means of the three components of the ILM measure make it clear that changes in the extent to which jobs are protected is a consistently important factor in the ILM change. Not much variation exists in

Table 1b. Means of Job-Relevant Variables by Quartiles of Firm Size, for Females, Private Sector

	Firm Size			
	Small	2	3	Large
Earnings	10,164	13,029	12,844	14,124
Education	13.3	13.1	13.5	13.1
Experience	12.4	13.7	13.7	13.3
Internal	1.02	1.79	2.49	2.32
Protection	.35	.61	.85	1.14
Up-mobility	.47	.73	1.01	1.18
Ladder	.20	.45	.63	.60
Reading	2.8	3.1	3.3	3.0
Writing	2.7	2.8	3.1	2.6
Union	.08	.15	.20	.25
Core	.44	.46	.52	.65
Authority	.71	.63	.55	.57
Supervise others	.44	.40	.37	.36
Say in hiring/firing	.27	.22	.19	.22
Formalization	2.5	3.9	4.2	4.3
Education required	.44	.48	.44	.33
Experience required	.52	.52	.48	.41
Black	.14	.16	.25	.15

the average degree that reading and writing is an important aspect of jobs, but the relationship seems to be curvilinear, with middle-sized firms most likely to house jobs whose performance requires more than usual amounts of reading and writing. The chances of being in a core industry or union job increases with size of firm, as expected, but males are in unions in substantially larger numbers only in the largest firms. Formalization increases with size, but the important increase comes when one moves from the smallest firms (<106) into the next category. Finally, contrary to assumptions in earlier work, there exists no straightforward positive relation between size and the requirement of either educational credentials or prior experience for hiring—the frequency of such requirements decreases in the largest firms.

Organizational characteristics that impinge on jobs seem to vary systematically by size, but not all linearly or even substantially. Further, the impact of these organizational size-related job characteristics on individual earnings has often been assumed or inferred, but rarely demonstrated. We turn to regression analysis to delineate these effects, regressing log(earnings) on log(firm size), log(establishment size), education, experience, hours worked, and race, then adding the most basic size-related variables: dummy variables for core/periphery and union job, the internal labor market measure, and the reading and writing requirement variables.[5] We use log(earnings) as the dependent variable rather than earnings, both because this is the usual and distributionally appropriate formulation and because it is easier to fit a model to the distribution after logging (the increase in R^2 is substantial).[6] We include for

---

[5] The only other variable covarying with size—formalization—was omitted after preliminary regressions showed it to have no effect on earnings for any individual category.

[6] We are generally sympathetic to Hodson's (1985) argument in favor of using unlogged earnings and his demonstration of the possible interpretive perils of not doing so when comparing gender or other groups with substantially different means (since in such cases the "rate of return" comparisons are confounded by different baselines, i.e., different returns). However, in this instance the choice of dependent variable is irrelevant to the outcome for two reasons. Substantively, we are not interested in comparing male and female rates of return (or return) to size, but rather the differences in the within-group patterning of such returns. Empirically, it turns out not to matter. The substantive patterns of

control purposes a dummy variable scored 1 for employment in a single-establishment firm. In our sample, 21.1 percent are employed in such firms and, as others have argued, it is a relevant characteristic when the effect of firm size is at issue (see Granovetter 1984; Hodson 1983; Hendricks 1976; Fuchs 1961).

Findings in columns 1 and 4 of Table 2 support the first hypothesis: net of education, experience, and number of hours worked, organization size has a significant positive effect on log(earnings) for both males and females. For males, firm size has the significant effect; for females, establishment size. The most general form of hypothesis 2 is also supported, as is evident in columns 2, 3, 5 and 6. Controlling for industrial sector, internal labor market, literacy requirements, and unionization removes the effect of size for males and reduces it for females (as well as increasing the explained variance). In columns 3 and 6, protection seems to be the most important component of the aspects of internal labor markets that we have measured, an outcome expected from the distribution of means in Table 1.

The findings are most clear-cut for males. Data in columns 7–12 of table 2 support hypotheses concerning size effects for white- and blue-collar workers (1a and 1b), but there is only partial support for the hypotheses concerning the mediation of those effects (2a and 2b) and the impact of job protection and mobility assurances (4a and 4b). Firm size, but not establishment size, has a significant positive effect on the earnings of white-collar males (column 7, hypothesis 1a), and this effect is not removed by controlling for individually salient organizational and industrial characteristics (columns 8 and 9, hypothesis 2b). Establishment size, not firm size, has a significant positive effect on the earnings of blue-collar males (column 10, hypothesis 1b), but this effect is also not removed by individually salient organizational and industrial characteristics (columns 11 and 12, refuting hypothesis 2a). When these equations are run for white males only, both parts of hypothesis 2 are unequivocally supported; that is, firm size is strongly positive and remains so for white-collar workers, while establishment size is strongly

---

findings in Tables 2 and 3 are duplicated when unlogged earnings are used instead as the dependent variable.

Table 2. Effect of Firm Size, Establishment Size, and Other Job Characteristics on Log (earnings), by Gender and Type of Job

	Males				Females		Males White Collar			Males Blue Collar		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log (Firm size)	.023**	.012	.010	.014	.009	.009	.035***	.024*	.023*	-.001	-.007	-.009
Log (Estab. size)	.014	-.002	-.003	.044***	.034***	.033***	-.016	-.028*	-.029*	.053***	.038***	.037**
Single-establishment	.005	.028	.016	-.023	-.007	-.010	.048	.065	.069	-.057	-.014	-.034
Education	.048***	.048***	.050***	.069***	.050***	.050***	.056***	.040***	.041***	.043***	.047***	.049***
Experience	.020***	.018***	.017***	.015***	.014***	.013***	.028***	.025***	.024***	.011***	.011***	.010***
Hours worked	.000***	.000***	.000***	.001***	.001***	.001***	.000***	.000***	.000***	.001***	.000***	.000***
Black	-.316***	-.352***	-.400***	-.122*	-.129***	-.097	-.344***	-.284***	-.269***	-.255***	-.309***	-.298***
Core		.202	.200***		.141***	.138***		.225***	.238***		.124**	.101*
Internal		.036***			.029*			.039***			.018	
Protection			.086***			.096***			.080**			.087***
Up-mobility			.011			-.027			.004			-.010
Ladder			-.027			.013			.023			-.084*
Reading		.027*	.026**		.046***	.049***		.029	.026		.020	.021
Writing		.026**	.026**		.036***	.033**		.040**	.042**		.012	.010
Union		.193	.181***		.032	.016		.024	.003		.275***	.267***
R^2	.46	.51	.51	.46	.50	.51	.49	.54	.54	.43	.49	.51

* $p < .10$.
** $p < .05$.
*** $p < .01$.

Table 2. Continued

	Females White Collar			Females Blue Collar			Females Clerical		
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Log (Firm size)	.016	.018	.018	.008	-.009	-.011	.036**	.036**	.032**
Log (Estab. size)	.029	.017	.015	.099***	.081***	.080***	-.014	-.006	-.007
Single-establishment	.058	.078	.069	-.147	-.174*	-.174*	-.022	.017	.007
Education	.069***	.055***	.055***	.035**	.025	.024	.020	.019	.018
Experience	.018***	.016***	.015***	.008*	.007*	.006	.013***	.013***	.012***
Hours worked	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***
Black	-.186**	-.190***	-.163**	.040	.006	.055	.041	.021	.084
Core		.173***	.168***		-.013	-.038		.089	.090
Internal		.020			.072**		.041	-.017	
Protection			.096***			.112**			.062
Up-mobility			-.044			.014			-.066*
Ladder			-.002			.126			-.094
Reading		.052***	.052***		.037	.048*		.006	.010
Writing		.040**	.038**		.013	.007		.034*	.029
Union		-.058	-.074		.086	.071		.162	.160
R^2	.46	.51	.52	.50	.55	.55	.51	.53	.55

* $p<.10$.
** $p<.05$.
*** $p<.01$.

positive (but controls remove the effect) for blue-collar workers. This suggests that the size effect is primarily indirect for white blue-collar males (as predicted), but not for black blue-collar males. Perhaps for black blue-collar workers an additional component of size is public scrutiny of large establishments (see the EEOC argument advanced below about female clerical workers). Job protection or sheltering has a significant positive impact on earnings for both groups of males (columns 9 and 12, hypothesis 4a). In columns 9 and 10 for white-collar males, establishment size becomes significant (at $p<.10$) in the presence of controls, but it has a significantly negative effect on earnings, net of firm size.[7] This is reasonable and expected finding, since it just means that while white collar males are certainly better off in larger firms, within those firms they benefit from being in smaller establishments. One obvious explanation is that the headquarters company of large firms is very often smaller than many of the firm's other establishments.

Hypothesis 4b, on the effect of mobility assurances, receives only slight support among males. While one of the mobility opportunity measures (ladder) is negative for blue-collar workers, as predicted, the other (up-mobility) is not, and neither has a net effect on earnings for white-collar workers. The latter finding could be due to white-collar opportunity chains being more informal and thus not tied to employer costs. This argument should not apply as directly to blue-collar workers (since within an establishment even the informal mobility avenues of blue-collar workers should be well known to management and taken into account in setting wage levels), making the absence of a significant negative

effect for up-mobility an unanticipated outcome. It nonetheless supports hypothesis 4b, that the only mobility measure that has an effect (and the predicted negative effect as well) is the measure of whether the employer had a formal mobility ladder.

In Table 3, columns 1 and 5 for white-collar males, firm size has a significant positive net effect in both nonmanufacturing and manufacturing industries. For blue-collar males, neither establishment size nor firm size has a significant effect on earnings in nonmanufacturing industries, but establishment size has a very strong positive effect in manufacturing industries (see columns 11 and 15). This effect is virtually undiminished by controls for organizational and industrial variables. These findings offer complete support for hypothesis 3. They also verify that Granovetter's findings concerning the paramount importance of establishment size apply only to blue-collar workers. [8] The results for females are somewhat less orderly. Data in column 4 of Table 2 reveal that establishment size has a significant positive net effect on log(earnings) for females, supporting hypothesis 1. When female workers are separated into white- and blue-collar categories (columns 13–18), the results are mixed. As predicted, establishment size is important to blue-collar workers (supporting hypothesis 1b), but firm size has no effect on the earnings of white-collar females (failing to support hypothesis 1a).[9] The effect of estab-

---

[7] That establishment size has a negative effect only net of firm size is an important distinction. Either will serve as a blanket proxy for organizational size, but not as a proxy for the other. That is, if only firm size was entered into the equations represented by columns 7 and 10 of Table 2, it would be significantly positive in both columns; the same is true for establishment size. Omitting either from an equation would seem to cause the effect of the one omitted to be collected into the apparent effect of the one used, masking the individual effects of each. Previous researchers who have used establishment size as a proxy for firm size have probably produced generalizations directly applicable to neither, but rather to some general size dimension. This finding holds (as do all findings reported in this paper) even when we remove all single-establishment firms (for which firm size equals establishment size) from the analysis.

[8] One finding for males does not precisely fit the argument. In columns 5 and 6 of Table 3 for white-collar males in manufacturing, the coefficient for the single establishment is large and statistically significant. Net of size, white-collar males in manufacturing seem to be better off in single-establishment firms. This is contrary to Granovetter's notion that outcome returns to size for white-collar workers result from the opportunity for mobility to component parts of a large firm. For higher male manufacturing workers, large, single-establishment firms are best. This finding does not fit our broader argument, for the positive effect of being in a single-establishment manufacturing firm is net of the size of that firm.

[9] For only non-clerical white-collar females, firm size still has no effect. Establishment size has some effect ($b=.05$, $p<.10$), and its effect is removed by ISOC controls. When only white females are considered, firm size does have a significant effect ($p<.10$) that is not removed by controls. However, firm size has no effect for non-clerical white-collar females. The effect of firm size for white-collar white females is apparently caused by the presence of clerical workers. Even for white females, then, hypothesis 1a is supported in fact, but not in spirit.

Table 3. Effect of Firm Size, Establishment Size, and Other Job Characteristics on Log (earnings), for Manufacturing and Nonmanufacturing Industries, by Gender and Type of Job

| | Males Nonmanufacturing | | | | Males Manufacturing | | | |
| | White Collar | | Blue Collar | | White Collar | | Blue Collar | |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log (Firm size)	.036*	.022	-.022	-.009	.044***	.035*	.023	.016
Log (Estab. size)	-.030	-.049**	.034	-.004	.010	.009	.090***	.083***
Single-establishment	.008	.032	-.170	-.062	.396**	.352**	.075	.055
Education	.059***	.043**	.055***	.056**	.060**	.051**	.033	.032**
Experience	.029***	.026***	.011***	.009**	.022***	.021***	.012***	.011***
Hours worked	.001***	.001***	.001***	.001***	.000***	.000***	.000***	.000***
Black	-.299**	-.276*	-.260***	-.244***	-.473**	-.320	-.178***	-.205
Core		.218***		.138*		.107		.118
Protection		.073		.107***		.081*		.067*
Up-mobility		.013		-.019		.010		-.004
Ladder		.091		-.139**		-.064		-.023
Reading		.019		.018		.029		.039**
Writing		.052**		.020		.023		-.016
Union		.078		.339***		-.110		.072
R^2	.50	.56	.46	.57	.48	.51	.51	.54

* $p<.10$.
** $p<.05$.
*** $p<.01$.

Table 3. Continued

| | Females Nonmanufacturing | | | | Females Manufacturing | | | |
| | White Collar | | Blue Collar | | White Collar | | Blue Collar | |
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Log (Firm size)	.015	.017	.007	−.080**	.013	.010	−.005	−.002
Log (Estab. size)	.032	.018	.094**	.112**	.020	.020	.106***	.106***
Single-establishment	.076	.058	−.124	−.386**	.113	.151	−.216**	−.208*
Education	.071***	.055***	.073**	.051	.071**	.063*	.022	.014
Experience	.021***	.018***	.008	.003	.009	.011	.010***	.010**
Hours worked	.001***	.001***	.001***	.001***	.001***	.000***	.000***	.000***
Black	−.178**	−.150*	.137	.122	−.021	−.093	−.053	−.057
Core		.151**		.213		.013		.058
Protection		.118***		.226**		.014		.012
Up-mobility		−.074**		.028		.034		.034
Ladder		.021		.380**		−.085		−.063
Reading		.063***		.067		.026		.012
Writing		.030		−.026		.042		.012
Union		−.012		.088		.103		−.036
R^2	.48	.55	.52	.62	.37	.42	.56	.58

* $p<.10$.
** $p<.05$.
*** $p<.01$.

lishment size on the earnings of blue-collar females is only slightly reduced by controls (contradicting hypothesis 2a). As expected, the effect of protection is strongly positive for both groups, as predicted by hypothesis 4a, but there is no negative effect for mobility assurances as predicted by hypothesis 4b.[10] It is also worth noting that being in a single-establishment firm has a strong negative impact on blue-collar female earnings; this is true in both manufacturing and nonmanufacturing firms (see Table 3). It can also be seen in Table 3 that hypothesis 3 does not seem to apply to females.

For clerical females separately, two other hypotheses are supported. Firm size, not establishment size, has significant positive impact (column 19, supporting hypothesis 1c), and this effect is not removed by controls for organizational and industrial attributes (columns 20 and 21, supporting hypothesis 2c). Protection has a significant positive effect for all females (column 6), corroborating hypothesis 4a, and upward mobility opportunities have a significant negative effect for clerical workers (column 21), offering further support for hypothesis 4b. Education does not significantly affect the earnings of clerical workers; only experience matters.

Earlier work on the individual-level impact of organization size emphasized not only superior outcomes for those in larger firms but also a superior earnings return to individual resources. Stolzenberg's (1978) entire focus was on return to education. Though we have addressed that to some extent by our replications and our examination of the effect of size net of education and experience, it is of interest to examine the effect directly. We regress log(earnings) on human capital variables (education, experience, education × experience, and experience-squared) for males and females, then calcu-

late the partial derivative of earnings with respect to education for each group. We evaluate this partial derivative for each individual at his or her level of the independent variable and then use it as a dependent variable in a second equation, regressing this "return to education" first on firm size and establishment size, then on those plus core, hours worked, protection, up-mobility, ladder, reading, writing, and union. This determines two things: 1) does organization size have a significant influence on individual earnings return to education; and 2) if so, is this influence mediated by organizational variables.

In Table 4 the patterning of the coefficients of log(firm size) and log(establishment size), when the dependent variable is the earnings return to education rather than log(earnings), differs from the earlier results. For both white-collar and blue-collar males, firm size has impact on earnings return to education. The effect of establishment size is also significant for white-collar males (though the latter is not true when only white males are considered). For both groups of females, establishment size has a significant effect.[11] This finding for males directly supports the earlier findings of Stolzenberg (1978), who argued for firm size differences in the individual income return to education. It also supports the contention of Kalleberg et al. (1981), who argued that Stolzenberg's findings apply only to white males. However, it contradicts the claim of Kalleberg et al. that females have a superior return to education only in small establishments; our data show superior returns for females in larger establishments. Note, however, our finding of inferior returns to education for females in large firms. Since Kalleberg et al. did not have both measures available, this fact could explain the difference in our findings (see note 7). These findings are not at variance with any of our hypotheses, since those hypotheses did not concern earnings return to education, but rather earnings net of education and experience. The two are quite different phenomena, having different theoret-

---

[10] When these equations are run for whites only, the effect of up-mobility is significantly negative, as predicted by hypothesis 4b. However, the effect of ladder is strongly positive for white blue-collar females (as it is for all females in nonmanufacturing industries). This not only refutes hypothesis 4b, it is also directly opposite to the finding for blue-collar males. Blue-collar males seem to lose earnings from the presence of formal job ladders, as expected, but white blue-collar females seem to gain earnings from the same. It is possible that for females, the wage-mobility "trade-off" suggested by hypothesis 4 is outweighed by the antidiscrimination impact of formal ladders.

[11] For clerical females, neither type of size has an effect on return to education (a result that is not surprising when one recalls the absence of any effect of education on the earnings of clerical workers). If clerical workers are removed, the coefficients for nonclerical white-collar workers are virtually unchanged in Table 4.

Table 4. Earnings Return to Education: Coefficients of Log Size ($\partial\$/\partial Ed$ dependent)

	Males				Females					
	White Collar Equation:		Blue Collar Equation:		White Collar Equation:		White Collar Non-Clerical Equation:		Blue Collar Equation:	
	1	2	1	2	1	2	1	2	1	2
log (firm size)	.0004***	.0003	.001*	.001*	-.001*	-.001	-.001**	-.001	.0000	.0000
log (estab. size)	.001**	.001*	-.001	-.001	.002***	.001**	.002***	.001*	.002*	.002*
R^2	.09	.17	.04	.05	.03	.08	.06	.14	.07	.19

Equation 1: $\partial\$/\partial Ed$ = f(Log [firm size], Log [estab. size], single establishment, black)
Equation 2: adds (Core/periphery, Protection, Up-mobility, Ladder, Reading, Writing, Union) to right-hand side

* $p<.10$.
** $p<.05$.
*** $p<.01$.

ical import and requiring different interpretations.

DISCUSSION

Using data from a large two-tier survey of workers and their employers, we have established that the employer-size effect on individual outcomes is partially a proxy for firm- and establishment-level attributes and industrial-sector attributes. Most importantly, we have shown that the effect is not mediated under all circumstances, and it is not uniform; indeed, it cannot be expected to exist under all circumstances. The effect of size is apparently mediated by the existence or lack of some form of internal labor market or union, by the importance of literacy skills to the job itself, by the location of the firm in the core or peripheral economic sector, but most importantly by the type of job at issue.

Most of our hypotheses concerning expected differences in the effect of firm and establishment size on earnings were supported for males, and fewer, but still most, of them were supported for females. One dimension underlying all of the hypotheses could be the extent to which persons, in their jobs, are "place-bound." That is, to the extent that a person's work career tends to be linked to a specific organization or suborganization, the attributes of the organization will have a stronger effect on that person's earnings. This idea is not too different from the old distinction between "job" and "career." The more place-bound one is, the more important the long-term salience of the particular place. This notion helps explain why establishment size should be more important to blue-collar workers than to white-collar workers, more important to females than to males (since the former are often place-bound for non-career reasons), and even why establishment size should not be important to clerical workers (since they have short advancement ladders, highly transferable skills, and exist in a high-demand market, they are not at all place-bound organizationally even when they might be geograhically). Whether this dimension is the key or not, our findings clearly indicate the need for some new organizing dimensions; existing models of organizational effects on individual earnings are too simple to describe a complex reality. A proper focus is not on whether individual effects, organizational effects, or industrial effects are pri-

mary, but rather on the conditions under which each of these very important aspects of economic life is most salient and on the conditions under which they interact differently.

Our hypotheses concerning the effects of size on female clerical workers were strongly supported. The hypotheses were developed with Kanter's descriptive work in mind, and based on the assumption that clerical mobility was intimately linked to managerial mobility on a one-to-one basis. This is probably an oversimplification, and is not the only possible explanation for our findings, as we indicated by suggesting the possibility of an internal-equity mechanism. Most clerical positions are essentially dead-end jobs. while there may be a ladder of income-step increases, there is rarely a well-defined progression of ever more responsible and prestigious jobs. Even when there is, the progression from clerk-typist to administrative assistant is both a relatively short ladder, and one not expected to be climbed by most in the occupation. It could well be that large firms, with their more extensive bureaucratic procedures, record-keeping, and paperwork, and thus more pressing need for efficient clerical workers, simply pay more to get them and keep them. This argument is weakened but not ruled out by the fact that controls for reading and writing requirements did not sufficiently diminish the firm effect; literacy is one component of higher worker quality, but there are many other components as well. That this interpretation is at least plausible is suggested by the statistically significant negative coefficient for the up-mobility variable for clerical females (see column 21 of Table 2). The sign of this coefficient means that a job with a high likelihood of being a stepping-stone to advancement is economically disadvantageous to females, a conceivable fact for most clerical jobs. Clerical job chains seem to be structured in a manner that would assure maximum retention of clerical workers by providing a ceiling to positional advancement and offering only monetary advancement; for clerical workers who will remain such, dead-end jobs are the best. Large firms, probably because of their position in the market, can offer the most in this regard.

Another possibility is that larger firms pay more to clerical workers because such workers are a large, easily observable pool of

females whose wages could come under EEOC scrutiny (see Bridges and Villemez 1986b for evidence that adherence to EEOC guidelines is directly related to organization size). A less cynical view would be that larger organizations better address comparable worth and gender pay-equity issues for their most visible set of females, whether for strategic or altruistic reasons.

All of these variants reduce to maintaining that clerical workers earn more in larger organizations simply because larger organizations choose to pay them more, for a variety of reasons. Though relying on a sponsored-mobility explanation is more theoretically tidy, this more obvious explanation could well be the primary one.

Our arguments concerning the different components of internal labor markets point to the need for more intensive investigation of this aspect of the workplace. One of the arguments—that sheltered jobs are more highly paid jobs—seems the most straightforward and follows directly from most previous literature. Its strong support here should not be controversial. The other argument, summarized in hypothesis 4b, will seem counterintuitive to many; indeed, it is accepted in much of the stratification and organization literature that the presence of formal (and informal) job ladders is beneficial to those on them. Our argument, drawn from a straightforward extension of neoclassical principles, was that such mobility structures could well be paid for by lower earnings (they could still be quite beneficial despite this). We did not find complete support for this hypothesis, but we found enough so that it still seems reasonable. We reported 15 equations containing the ladder and up-mobility variables (7 and 8 each in Tables 2 and 3). In only five of these was the coefficient of either variable statistically discernible from zero. Of those five, the coefficients were negative in four cases (as predicted) and positive in only one. The positive coefficient was for the formal ladder variable for blue-collar females. Though we offered a possible explanation for this case (that formal ladders, though detrimental to earnings, are far less detrimental than the informal gender discrimination they replace), it is clear that the specific hypothesis requires elaboration: it does not seem to be correct as stated. What does receive support is our conception of internal labor markets as congeries of different aspects, some directly

beneficial to workers, some a trade-off, and some detrimental. This complex feature of the workplace deserves far more study than it has received. Such study is not easily accomplished; those few excellent studies that do have detailed information on specific organizational ladders can only speculate as to what outcomes would be if they were not in place.

In summary, establishment size and firm size are quite different entities, each with its own effect and each uniquely affecting different portions of the workforce. Organizational attributes have different individual effects in varying industrial contexts, and differing organizational size-related internal attributes have different effects on various occupational groups. We have presented findings that begin to disentangle these effects.

REFERENCES

Addison, John T. and W. Stanley Siebert. 1979. *The Market For Labor: An Analytical Treatment.* Santa Monica, CA: Goodyear Publishing Co.

Althauser, Robert P. and Arne L. Kalleberg. 1981. "Firms, Occupations, and the Structure of Labor Markets." Pp. 119–49 in *Sociological Perspectives on Labor Markets,* edited by Ivar Berg. New York: Academic Press.

Bailey, William R. 1980. "Wage Rate Variation by Size of Establishment." Industrial Relations 19 192–98.

Baron, James N. 1984. "Organizational Perspectives on Stratification." Pp. 37–70 in *Annual Review of Sociology,* vol. 10, edited by Ralph H. Turner. Palo Alto, CA: Annual Reviews.

Baron, James N. and William T. Bielby. 1980. "Bringing the Firms Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737–65.

———. 1984. "The Organization of Work in a Segmented Economy." *American Sociological Review* 49:454–73.

Baron, James N., Alison Davis-Blake, and William T. Bielby. 1986. "The Structure of Opportunity: How Promotion Ladders Vary Within and Among Organizations." *Administrative Science Quarterly* 31:248–73.

Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics,* 3:33–58.

Bridges, William P. and Wayne J. Villemez. 1986a. "Informal Hiring and Income in the Labor Market." *American Sociological Review* 51:574–82.

———. 1986b. "Overeducated Minority Workers: Does EEO Coverage Make a Difference?" Pp. 72–90 in *Affirmative Action: Theory, Analysis, and Prospects,* by Michael W. Combs and John Gruhl. Jefferson, NC: McFarland.

Burawoy, Michael. 1979. *Manufacturing Consent: Changes in the Labor Process Under Monopoly Capitalism.* Chicago: University of Chicago Press.

Child, John. 1973. "Predicting and Understanding Organization Structure." *Administrative Science Quarterly* 19:168–85.

Cole, Robert E. 1979. *Work, Mobility, and Participation: A Comparative Study of American and Japanese Industry*. Berkeley: University of California Press.

Fuchs, Victor R. 1961. "Integration, Concentration, and Profits in Manufacturing Industries." *Quarterly Journal of Economics* 2: 278–91.

Granovetter, Mark. 1984. "Small is Bountiful: Labor Markets and Establishment Size." *American Sociological Review* 49:323–34.

Hall, Richard H. 1982. *Organizations: Structure and Process*. Englewood Cliffs, NJ: Prentice-Hall.

Hendricks, Wallace. 1976. "Conglomerate Mergers and Collective Bargaining." *Industrial Relations* 15:75–87.

Hodson, Randy. 1985. "Some Considerations Concerning the Functional Form of Earnings." *Social Science Research* 14:374–94.

Hodson, Randy. 1983. *Workers' Earnings and Corporate Economic Structure*. New York: Academic Press.

———. 1984. "Companies, Industries, and the Measure of Economic Segmentation." *American Sociological Review* 49:335–48.

Kalleberg, Arne, Michael Wallace, and Robert Al-

thauser. 1981. "Economic Segmentation, Worker Power and Income Inequality." *American Journal of Sociology* 87:651–83.

Kanter, Rosabeth Moss. 1977. *Men and Women of the Corporation*. New York: Basic Books.

Lester, R. 1967. "Pay Differentials by Size of Establishment." *Industrial Relations* 7:57–67.

Masters, S.H. 1969. "An Interindustry Analysis of Wages and Plant Size." *Review of Economics and Statistics* 51:341–45.

Oi, Walter Y. 1983. "Heterogeneous Firms and the Organization of Production." *Economic Inquiry*, 21:147–71.

Phelps Brown, H. 1977. *The Inequality of Pay*. Berkeley: University of California Press.

Podgursky, Michael. 1986. "Unions, Establishment Size, and Intra-Industry Threat Effects." *Industrial and Labor Relations Review* 39:277–84.

Stolzenberg, Ross. 1978. "Bringing the Boss Back In." *American Sociological Review* 43:813–28.

Tolbert, Charles M., Patrick Horan, and E.M. Beck. 1980. "The Structure of Economic Segmentation: A Dual Economy Approach." *American Journal of Sociology* 85:1095–16.

BI-ETHNIC LABOR MARKETS, MONO-ETHNIC LABOR MARKETS, AND SOCIOECONOMIC INEQUALITY*

MOSHE SEMYONOV

Tel Aviv University and *University of Illinois at Chicago*

The research examines the role of local opportunity structure in the determination of social and economic outcomes by focusing on the impact of local labor market segregation on socioeconomic inequality between Jews and Arabs in Israel. Analysis of the 1983 Israeli Census reveals that the rules by which minority members achieve socioeconomic status vary considerably across bi-ethnic and mono-ethnic labor markets. Arabs working in Jewish communities (bi-ethnic labor markets) suffer the detrimental consequences of both occupational and income discrimination more than any other group. By contrast, Arabs working in Arab communities (mono-ethnic labor markets) are occupationally advantaged. With the absence of competition, they produce incumbents for positions usually held by Jews. The findings suggest that the magnitude of disadvantages that superordinate groups force on minority groups is highly dependent on the degree of group competition in the local labor market.

Students of social stratification have traditionally used individual-level variables within status-attainment models to examine the distribution of social and economic rewards. Recently, more researchers have advanced structural explanations for understanding socioeconomic and ethnic inequalities (e.g., Baron and Bielby 1980; Bibb and Form 1977; Hudson and Kaufman 1982; Kaufman 1984; Spilerman 1977; Stolzenberg 1975). Among the structural explanations, however, only a few have stressed the local opportunity structure as a major determinant of social and economic outcomes (e.g., Spilerman and Habib 1976; Parcel 1979; Parcel and Mueller 1983; Logan 1976, 1978; Semyonov 1981; Lewin-Epstein 1985, 1986). According to this literature, one reason why individuals are differentially successful is because they live in places with different opportunity structures. Places are viewed as labor markets, an independent dimension of the stratification system, exerting significant influence on socioeconomic and ethnic inequalities. This paper expands our knowledge of the role of local labor markets in the distribution of social and economic rewards. It focuses on the impact of labor-market segregation on occupational and economic inequality between Arabs and Jews in Israel.

THEORETICAL CONSIDERATIONS

Spatial segregation has long been viewed by students of social stratification and race relations as an effective mechanism through which minorities are denied access to opportunities and rewards, and thus, as one of the major determinants of ethnic inequalities. In a relatively free economy, however, a minority member can avoid the detrimental consequences of segregation by changing place of residence (or place of work). Integration of a minority member into another labor market appears to be rather costly. A large number of studies have repeatedly demonstrated that the socioeconomic disadvantages (mostly occupational) of minorities (mostly blacks, in America) tend to rise as their relative size in the community population increases (e.g., Fossett 1984; Semyonov, Hoyt, and Scott 1984; Frisbie and Neidert 1977; Glenn 1964, 1966; LaGory and Magnani 1979; Martin and Poston 1972; Wilcox and Roof 1978; Tienda and Lii 1987).

Two complementary arguments have been advanced to explain the relationship between the rise in minority population and growth in social and economic differentials. The first suggests that an increase in the relative size of a minority population generates a threat to

superordinates based on greater job competition (e.g., Williams 1947; Allport 1954; Blalock 1967; Bonacich 1972, 1976). Antagonism, hostility, and motivation to discriminate grow as the proportion of a subordinate population in the community rises. That is, minority members become increasingly economically depressed as they become increasingly numerous potential competitors for the jobs of majority members. To be competitors, the two groups must enjoy some degree of integration. Undoubtedly, they must work, or must have the potential to work, in the same local labor market.

The other view contends that an influx of subordinate population increases the supply of cheap labor as a target for economic exploitation. Since occupations are split across racial (or ethnic) lines, the availability of minority members increases the potential pool of candidates to fill the least desirable, low-pay jobs (e.g., Glenn 1964; Spilerman and Miller 1977; Semyonov et al. 1984). The presence of a large minority population thus enables superordinates to abandon the low-status jobs and "flow" to more prestigious and lucrative occupations. Consequently, occupational and economic differentiation between minority and majority members tend to widen as the proportion of the former in the population increases (e.g., Glenn 1964; Spilerman and Miller 1976; Semyonov et al. 1984). Furthermore, as the relative size of a minority within a local labor market rises, so does the probability that any two randomly selected workers from that labor market will be of different ethnicity; hence, the advantages of discrimination for employers also increase (Thurow 1969). In light of these arguments, it may be suggested that as integration into a bi-ethnic labor market increases, so do the relative disadvantages of a minority population.

The concept *integration* (or *segregation*) in this article differs from that implied by conventional concepts of integration as measured by indices of residential segregation within a city. No doubt, the latter sort of integration reduces the socioeconomic disadvantages of minorities. Integration, as employed here, is considered the probability that two randomly selected persons in the same local labor market will be of opposite ethnicity (minority versus majority) (cf. Semyonov and Tyree 1981). Since this phrase is too cumbersome for repeated use, the term

integration (or *segregation*) will be used instead, but its operational meaning should be kept in mind.

The logic of this framework suggests that, while segregation into ethnic enclaves (mono-ethnic market) excludes minorities from equal access to opportunities and rewards, it may provide them with temporary protection from discrimination generated by competition. That is, a mono-ethnic labor market can free subordinate members from direct competition with superordinates. As a result, minorities can enjoy job opportunities that otherwise would be denied to them and grasped by superordinates (Frazier 1951; Semyonov and Tyree 1981; Lewin-Epstein 1986).

In other words, when a minority population reaches a critical mass and is large enough to develop independent, mono-ethnic, labor markets, it may be able to mobilize resources and produce incumbents for positions usually held by members of the majority population (Frazier 1951; Fischer 1975; Semyonov and Tyree 1981; Lewin-Epstein 1986). This idea has been suggested in the ethnic enclave literature (e.g., Light 1972) and was clearly articulated by Lieberson (1980; pp. 297–98):

> Among other factors, as a group gets larger it is likely to develop certain internal strengths that will support some occupational activities even if outsiders are totally against their holding the position. Hence, if the black population base is large enough, there will be support for black doctors, black clergy and so on, even if they remain totally unacceptable to others. Likewise, there will develop certain entrepreneurial possibilities and other employment shifts will occur.

This argument does not mean that subordinates actually benefit from segregation, or that minorities reap absolute occupational advantages in a mono-ethnic labor market. They do not. Segregation is a structural device through which minorities are excluded from opportunities and rewards. The relative cost of discrimination however, is expected to be greater in bi-ethnic than in mono-ethnic labor markets.

THE SETTING

Israel is a particularly appropriate setting to examine the relationship between labor-market segregation and ethnic-linked economic inequalities. The two ethnic groups studied are Jews and non-Jews (mostly

Arabs). The distinction between Arabs and Jews is the most meaningful ethnic split in Israel. Arabs comprise about 15 percent of the population of Israel.[1] They are subordinate to Jews in every aspect of social and economic status, such as education, occupation, income, and political power. Their cultural orientation is more traditional, and most reside in villages and small towns (e.g., Peres 1971; Smooha 1978; Simon 1978; Semyonov and Tyree 1981). Furthermore, Arabs are highly segregated from Jews. Most communities in Israel are overwhelmingly Arab or Jewish. In fact, there are only seven communities in Israel that are formally defined by the Israeli Central Bureau of Statistics as mixed localities.[2] The value $D = 75$ for the index of dissimilarity for the distribution of Jews and Arabs across communities (reported by Semyonov and Tyree 1981) reflects the extreme segregation between the groups. Simply, this means that three-quarters of either population would have to change community of residence for the two groups to live in the same locality.

The spatial segregation between Jews and Arabs represents differential access to opportunities. Arabs are more likely than Jews to reside in small communities and in rural places, where occupational and economic opportunities are scarce (Peres 1971; Semyonov and Tyree 1981). Consequently, many Arabs in Israel search for jobs in Jewish communities. Thus, it is sociologically meaningful to distinguish among three types of Arab populations in the Israeli labor force: those who live and work in Arab communities (hereafter "segregated"), those who live in Arab localities but work in Jewish places (hereafter "commuters"), and those who live

and work in Jewish or mixed places (hereafter "residents").

The size of the disadvantage a superordinate group can force on a minority group is a function of the degree of competition between the groups. Integration, to the extent of living and working in the same local labor market, is a conducive, if not necessary, condition to assure such competition. Following this logic, we expect that Arabs living or working in Jewish communities (bi-ethnic labor markets) suffer the detrimental consequences of socioeconomic discrimination more than Arabs living and working in Arab places (mono-ethnic labor markets). More specifically, we expect that residents will be most disadvantaged; that segregated, who are free of direct competition with Jews, will be least disadvantaged, if not relatively advantaged; and that the commuters will fall between these two groups.

ANALYSIS AND FINDINGS

Data and Variables

Data for the present investigation were obtained from the 20 percent sample of the Israeli Census conducted by the Israel Central Bureau of Statistics in 1983. The census tape contains data on demographic and socioeconomic characteristics of the labor-force population as well as on place of residence and place of work.[3] The distinction between place of residence and place of work identifies the three groups of Arabs noted in the previous section: segregated, residents, and commuters. This distinction provides a unique opportunity to examine the impact of labor-market segregation on socioeconomic inequalities between minority and majority populations.

---

[1] It should be emphasized that Israeli Arabs are citizens of Israel. This study does not focus on noncitizen Arabs from the West Bank and Gaza Strip who find employment in Israel. From a legal viewpoint, every Israeli citizen, Jew or non-Jew, can work anywhere in Israel.

[2] There are 103 identifiable communities (size ranges from 5,000 to 400,000 residents) in the Israeli Census. The percent of Arab residents in the seven "mixed localities" ranges from approximately 6 to 30. All other places are either Arab (35) or Jewish (61). That is, there are virtually no Jews in Arab localities and almost no Arabs in Jewish places. Furthermore, while many Arabs commute to work in Jewish communities, hardly any Jews find employment in the smaller Arab communities. Thus, identifying specific "cutoffs" were irrelevant for this research.

---

[3] The census is conducted by the Israel Central Bureau of Statistics once every 11 years. It is a systematic random sample of households stratified within region of residence. Data were obtained for 1,151,000 households across 103 places (communities) of 5,000 residents or more. Detailed labor force and socioeconomic characteristics were obtained for one-fifth of the sample for all persons 15 years or over in the household. According to the Central Bureau estimates, Jews were slightly oversampled (20.5 percent) and Arabs were only slightly undersampled (19.98 percent) compared to 20 percent expected value (Central Bureau of Statistics, 1985). The analysis reported in this paper was restricted to labor force population 25–64 years of age.

A set of individual-level variables, traditionally employed in models of status attainment, were chosen for the analysis: gross monthly income (in Israel shekels), occupational status (measured on Tyree's [1981] 100-point scale of occupations in Israel), years of formal schooling, age (in years), hours of work per week, and gender. In the analysis that follows I shall first compare the characteristics of the various subpopulations (i.e., Jews, segregated, commuters, and residents); second, examine models of status attainment across mono-ethnic and bi-ethnic labor markets for the various subpopulations; and third, decompose the mean socioeconomic differential between subpopulations to estimate the sources of socioeconomic inequalities across bi-ethnic and mono-ethnic labor markets.

Descriptive Overview

The subpopulations are compared in Table 1, which includes the mean socioeconomic, demographic, and residential characteristics of Jews, segregated, residents, and commuters. The data demonstrate considerable differences among the groups with regard to size of place of residence. As a whole, Jews are more likely than Arabs to reside in large urban centers where occupational opportunities are abundant and less likely to dwell in rural and small localities where opportunities are scarce. The difference is most pronounced for commuters, who are heavily drawn from small places in search of employment opportunities. By contrast, residents tend to concentrate in large urban centers, more than any other population group.

The differences between Jews and Arabs are also evident with regard to socioeconomic characteristics. Arabs are inferior to Jews in every aspect of socioeconomic status included in Table 1: formal education, occupational status, and income. They are also considerably younger than Jews. These socioeconomic differentials are somewhat underestimated due to the greater participation of

Table 1. Characteristics: Means (Standard Deviations) and Category Percentages of Jews and Arab Population Groups of Israeli Labor Force Population, Aged 25–64, 1983

Variables	Jews Total (1)	Arabs Total (3+4+5)	Segregated (3)	Commuters (4)	Residents (5)
Monthly income	33,760.73 (34,835.56)	21,117.73 (15,083.40)	22,603.71 (14,871.16)	21,065.29 (13,815.33)	19,362.58 (16,198.36)
Occupational status	46.16 (19.08)	38.45 (19.01)	42.92 (21.35)	32.22 (14.85)	37.72 (18.77)
Years of schooling	12.0 (3.5)	9.9 (3.5)	10.3 (3.9)	8.7 (3.1)	10.0 (4.2)
Years of age	40.6 (11.17)	37.3 (9.7)	36.8 (9.5)	36.2 (9.4)	38.8 (10.2)
Hours of work per week	42.0 (14.2)	42.4 (12.4)	40.0 (13.5)	45.2 (10.3)	43.3 (12.5)
Percent in Major Occupational Categories					
Professional managers	32.0	22.1	32.0	9.0	20.0
Clerical and sales	27.0	16.2	17.0	10.0	20.0
Manual and service	41.0	61.7	51.0	81.0	60.0
Percent Women in Labor Force	40.0	17.06	20.0	7.00	21.00
Percent in Locality of Residence by Size					
Less than 10,000	14.0	40.8	60.0	62.0	2.0
10,000–20,000	7.0	14.6	20.0	23.0	2.0
20,000–50,000	18.0	16.5	20.0	14.0	14.0
50,000–100,000	10.0	0.3	—	—	1.0
100,000 and over	51.0	27.5	—	—	81.0
Population N	211,417	12,727	5,207	3,206	4,314

Jewish women than Arab women in the cash economy.[4]

The Arab population, however, cannot be treated as a homogeneous unit. There are considerable variations among the subgroups.[5] The commuters are the least educated group; apparently, they are unable to compete on equal terms with either Jews or Arabs. Consequently, they have to settle for the low-status, least desirable occupations relinquished by Jews. Their income, nonetheless, is relatively high. Perhaps they exchange occupational status for income. Segregated and residents acquired, relative to commuters, higher levels of education. The former, free from direct competition with Jews, were able to attain higher occupational status as well as higher incomes.

Especially illuminating are the results regarding the occupational structure of the various Arab populations. Both mean occupational status and standard deviation for segregated are considerably higher than for either commuters or residents. Segregated are able to produce, within their own independent labor markets, exactly the same proportion of professionals and mangers as Jews do. Commuters, however, supply mainly manual and service workers for the Jewish labor markets. Residents, despite their educational level, are somewhat underrepresented among all white-collar (professionals and managers, as well as clerks and sales) occupations, and somewhat overrepresented among manual and service workers.

The values in Table 1 support the theoretical expectations. Segregated, though inferior in status to Jews, are superior to the two other Arab populations. They appear to be somewhat protected from economic discrimination. By contrast, residents are the most disadvantaged group. Their occupational status is relatively low in comparison to their education, and their incomes are relatively low in comparison to their occupations. The commuters are between these two extremes.

Apparently, a selectivity bias produces somewhat different relations between human resources (e.g., education) and socioeconomic status (e.g., occupation, income) among commuters. The commuters tend to reside in smaller places where opportunities are scarce. Furthermore, their lower education and younger age hinder their ability to compete successfully, not only with Jews but also with other Arabs. Thus, they are compelled to take manual and service occupations available in Jewish communities and trade off occupational status with income.

Socioeconomic Attainment Across Labor Markets

Although interesting, the differences among the groups do not tell us about the relative impact of labor-market segregation on socioeconomic inequalities or whether socioeconomic status is differentially determined for minority members in mono-ethnic and bi-ethnic labor markets. That is, the descriptive data do not provide accurate estimates as to whether the process of status attainment varies across local labor markets. Thus, Table 2 includes coefficients of a pair of regression equations for each population group (Jews, segregated, residents, and commuters). In equation 1, occupational status is predicted as a function of education, age, and gender. In equation 2, monthly income is taken as a function of age, education, occupational status, hours of work, and gender.[6]

The regression coefficients in Table 2 represent the direct net effect of each independent variable included in the equation on occupational status (equation 1) or on income (equation 2). The results are virtually identical whether or not size of community of residence was included in the regression equation. Regardless of the population group considered, the most significant predictor of occupational status in equation 1 is education. That is, education is likely to improve occupational status. The status return on education is highest for segregated and Jews and lowest for commuters and residents. Age in equation 1 has positive impact on occupational status in all groups. Gender, however,

---

[4] A separate analysis for the male population resulted in conclusions identical to those reported in this paper.

[5] There is also some variation among Arabs on the basis of religious affiliation. The most meaningful one is the distinction between Christian, Moslem, and Druze. It should be emphasized, however, that separate analyses of these groups resulted in similar findings and patterns to those of the total populations. This separate analysis is not presented here for the sake of parsimony. Results are available from the author on request.

[6] In equation 1d and 2d, a set of dummy variables representing seven categories of community size was added to control for market conditions that may be associated with locality size.

Table 2. Regression Equations Predicting Occupational Status[a] (Eq. 1) and Log. Income[b] (Eq. 2) for Jews and Arab Population Groups of the Israeli Labor Force, Age 25–64, 1983

Independent Variables	Jews Occup. Status Equation 1		Jews Log. Income Equation 2		Arabs — Segregated Occup. Status Equation 1		Arabs — Segregated Log. Income Equation 2		Arabs — Commuters Occup. Status Equation 1		Arabs — Commuters Log. Income Equation 2		Arabs — Residents Occup. Status Equation 1		Arabs — Residents Log. Income Equation 2	
	1	1[d]	2	2[d]	1	1[d]	2	2[d]	1	1[d]	2	2[d]	1	1[d]	2	2[d]
Gender[c] b	-3.632	-3.616	-.379	-.379	4.037	4.036	-.357	-.358	-1.238	-1.237	-.402	-.402	.582	.610	-.245	-.319
s.e.	(.072)	(.072)	(.004)	(.004)	(.558)	(.558)	(.023)	(.023)	(.875)	(.875)	(.039)	(.039)	(.591)	(.599)	(.028)	(.027)
Years of age b	.079	.068	.005	.005	.201	.197	.012	.012	.117	.115	.002	.002	.078	.072	.002	.002
s.e.	(.003)	(.003)	(.0001)	(.0004)	(.025)	(.025)	(.0009)	(.0009)	(.025)	(.025)	(.001)	(.001)	(.023)	(.023)	(.001)	(.001)
Years of schooling b	3.590	3.516	.031	.032	3.989	3.985	.042	.042	2.846	2.845	.014	.014	2.921	2.912	.037	.033
s.e.	(.010)	(.010)	(.0006)	(.0006)	(.059)	(.059)	(.003)	(.003)	(.070)	(.070)	(.004)	(.004)	(.059)	(.059)	(.003)	(.003)
Hours of work per week b	—	—	.016	.016	—	—	.005	.005	—	—	.008	.008	—	—	.009	.009
s.e.	—	—	(.0001)	(.0001)	—	—	(.0007)	(.0007)	—	—	(.0009)	(.0009)	—	—	(.0008)	(.0008)
Occupational status b	—	—	.013	.013	—	—	.011	.011	—	—	.010	.010	—	—	.008	.009
s.e.	—	—	(.0001)	(.0001)	—	—	(.0006)	(.0006)	—	—	(.0008)	(.0008)	—	—	(.0007)	(.0007)
Constant	4.990	1.988	8.833	8.792	-10.386	-10.752	8.742	8.743	4.538	4.684	9.335	9.344	4.762	3.963	8.782	9.276
R^2	.41	.42	.38	.38	.52	.52	.32	.32	.39	.39	.17	.17	.41	.41	.19	.25

[a] Occupational status measured on Tyree's (1981) 100 point-scale of socioeconomic status of occupation in Israel.
[b] Income (converted in logarithmic scale) is gross monthly income in Israeli shekels.
[c] Men coded 1; Women 2.
[d] The coefficients reported in equations 1[d] and 2[d] are effects net of community size (see footnote 6 for details).

operates differently among the various groups. Jewish women are disadvantaged in the attainment of occupational status compared to Jewish men, while segregated women are advantaged over segregated men in the attainment of occupational status.[7] Gender differences in the attainment of occupational status among commuters and residents are negligible and statistically not significant.

Equation 2 reveals similar patterns in the determination of income among all population groups, regardless of the local labor market. The incomes of men are considerably higher than those of women. Education, occupational status, age, and hours of work all exert positive effects on income. That is, income is likely to increase with rising levels of these four independent variables. It is interesting to note, however, that the impact of education on income (as well as the coefficient of determination) is especially low for commuters. Perhaps education, as well as other human resources, are not as relevant for the income of commuters.

Decomposing Mean Differentials

The findings reported thus far provide only partial and indirect support for the argument that a bi-ethnic labor market produces greater disadvantages to the minority population, while a mono-ethnic labor market hinders discrimination generated by competition. This thesis, however, needs further analysis. The extent to which the various groups of Arabs gained or lost socioeconomic status due to human capital, ethnic membership, or the local labor market must be established. For example, it is not clear to what extent the lower income of residents is a function of their inferior occupational status, education, and age, or a function of the bi-ethnic labor market in which they operate. Decomposing mean socioeconomic differentials, first, between each group of Arabs and the Jewish population, and, second, between Arab population groups, can provide accurate estimates of the extent to which mono-ethnic or·

bi-ethnic labor markets are socioeconomically beneficial or detrimental to the minority population. The decomposition procedure will also provide accurate estimates of the sources of socioeconomic inequality across labor markets.

There are several parallel models for decomposing mean differences between groups via regression equations (e.g., Duncan 1968; Winsborough and Dickenson 1971; Althauser and Wigler 1972; Oaxaca 1973; Iams and Thornton 1975; Jones and Kelly 1984). These models explain economic discrimination by identifying several components that account for the socioeconomic differentials between two groups. The model adopted in this work for partitioning observed mean differences between two groups follows the logic and procedure recommended by Jones and Kelly (1984):

$$Y^H - Y^L = [(a^H + \Sigma b^H X^L)$$
$$- (a^L + \Sigma b^L X^L)]$$
$$+ \Sigma b^L (X^H - X^L) +$$
$$\Sigma (b^H - b^L) (X^H - X^L)$$

The Ys are the mean values of the dependent variables of the H (High Status) and L (Low Status) groups. The Xs are the mean values of the antecedent variables included in the equation, bs are the regression coefficients, and the as are the two respective intercepts.

The model identifies three components. The first component is the unexplained differences between the groups and is due to both group membership and returns on human resources. It can be regarded as representing ·"market discrimination." The second component is the portion of the gap due to differences in human resources (or composition or endowment). The third component is the interaction effect of jointly changing both mean resources and coefficients over the effect of changing them one at a time.[8]

---

[7] The relative occupational advantage of segregated women over segregated men results from the type of occupations available for women in the Arab enclave. In the traditional Arab culture, the few women who participate in the cash economy are mostly employed in white-collar and professional occupations. Many more men, however, are employed in low-status manual occupations.

[8] This decomposition model was chosen for both substantive and methodological reasons. First, it identifies three components that are most appropriate for the questions addressed in this study. Second, the use of the intercepts as a separate component may generate a problem in interpretation. The intercepts are derived from the regression coefficients. They have no clear interpretation in equations including variables with no meaningful zero point. Thus, following the recommendation made by Jones and Kelley, the "discrimination component" includes both returns and intercept. (For a detailed discussion of the methodological considerations regarding various models of decomposition, see Jones and

Table 3. Components of Occupational Status Differentials Between Pairs of Groups, Israeli Labor Force, Age 25–64, 1983

	Pair			Comparisons	
	(1) Jews vs. Residents	(2) Jews vs. Commuters	(3) Jews vs. Segregated	(4) Segregated vs. Commuters	(5) Segregated vs. Residents
Observed Gap					
$Y^H - Y^L$	8.439	13.941	3.240	10.702	5.199
(100%)					
Market Discrimination					
$(a^H + \Sigma b^H X^L) - (a^L + \Sigma b^L X^L)$	1.884	2.988	−2.402	3.710	4.508
(%)	22.0	21.0	−74.0	34.7	86.7
Composition Resources					
$\Sigma b^L (X^H - X^L)$	6.030	9.462	8.330	4.451	.675
(%)	71.0	68.0	257.0	41.6	13.0
Interaction					
$\Sigma (X^H - X^L)(b^H - b^L)$	.525	1.491	−2.688	2.541	.016
(%)	7.0	11.0	−83.0	23.7	0.3

Table 3 presents the results of the application of this decomposition procedure to equation 1. The first three columns pertain to the occupational gap between Jews and each Arab population. The last two columns pertain to the occupational differences between segregated and residents and segregated and commuters, respectively. The results described in this table lend firm support to the argument that Arabs living and working in Jewish places (bi-ethnic labor market) suffer the detrimental consequences of occupational discrimination more than Arabs working in Arab places (mono-ethnic market).

A considerable portion of the occupational gap between Jews and either commuters or residents (about 70 percent) can be attributable to the lower education and younger age of the Arab population. About 20 percent of this gap is viewed, according to the model, as resulting from market discrimination. The findings, however, are quite different for the segregated. When considering the lower education and comparative youth of this group, the observed gap between Jews and segregated would be expected to increase by more than five points. That is, the occupational status of segregated Arabs is actually higher than one would expect on the basis of their human resources. Subsequently the negative-discrimination component clearly indicates that segregated are occupationally

overpaid. Furthermore, the negative-interaction component indicates that the combination between resources and return on human resources is advantageous for Arabs in their ethnic enclave. Apparently, occupational attainment of Arabs is enhanced in a mono-ethnic labor market and hurt in the bi-ethnic labor market.

The results presented in columns 4 and 5 of Table 3 provide further evidence and reaffirm the argument that employment in a bi-ethnic labor market has detrimental consequences for the occupational status of the minority population. While over 40 percent of the occupational advantage of segregated over commuters is attributable to their superior human resources, most of their occupational advantages over residents is attributable to market discrimination.

Table 4 contains results obtained by applying the decomposition procedure to equation 2. Columns 1, 2, and 3 reveal that only a small portion of the income disparity between Jews and Arabs can be attributable to human resources. Even when considering the lower occupational status, lower educational levels, and younger age of Arabs, a considerable portion of the income gap remains attributable to ethnicity and differential returns on human resources (market discrimination). Consequently, the discrimination components account for 82, 77, and 61 percent of the income gaps between Jews and residents, segregated, and commuters, respectively. Apparently, had Arabs' earnings been determined similarly to Jews, their incomes could have been considerably higher.

Kelley [1984].) It should be emphasized, however, that other decomposition techniques (e.g., Duncan 1968; Iams and Thornton 1975) lead to the same conclusions reported in this paper. The findings, indeed, are not model specific.

Table 4. Components of Income Differentials Between Pairs of Groups, Israeli Labor Force Population, Age 25–64, 1983

	Pair			Comparisons	
	(1) Jews vs. Residents	(2) Jews vs. Commuters	(3) Jews vs. Segregated	(4) Segregated vs. Commuters	(5) Segregated vs. Residents
Observed Gap $Y^H - Y^L$ (100%)	.487	.332	.296	.035	.191
Market Discrimination $(a^H + \Sigma b^H X^L) - (a^L + \Sigma b^L X^L)$	.400	.203	.228	−.084	.159
(%)	82.0	61.0	77.0	−240.0	83.0
Composition Resources $\Sigma b^L (X^H - X^L)$	.091	.043	.089	.042	.024
(%)	19.0	13.0	30.0	120.0	13.0
Interaction $\Sigma (X^H - X^L)(b^H - b^L)$	−.004	.086	−.021	.077	.008
(%)	−1.0	26.0	−7.0	220.0	4.0

Columns 4 and 5 of Table 4 include the results obtained from decomposing mean differences between income of segregated and commuters, and segregated and residents, respectively. The results further demonstrate the differential effect of local labor-market segregation on the attainment of income. Only 13 percent of the income gap between segregated and residents is due to differences in human resources (especially occupational status). The remaining 83 percent of the gap is regarded as resulting from group membership and returns on human resources (discrimination component associated with market conditions). Commuters, by contrast, are overpaid compared to segregated. Given their low occupational status, low education, and younger age, commuters' incomes are higher than what they could possibly earn in their own labor markets. This relative advantage is reflected by the negative discrimination component in the table. Apparently, commuters trade off occupational status with income on joining the bi-ethnic labor market. They sacrifice occupational status, but compensate for this loss by earning the relatively higher incomes available in Jewish communities.

SUMMARY AND CONCLUSION

The major objective of this research was to investigate the role of local labor markets in the determination of social and economic outcomes in general, and to examine the impact of labor-market segregation on ethnic inequalities in particular. It was proposed that the extent of the disadvantages a superordi-nate group can force on a subordinate group is dependent in large measure on competition between the groups. Following this logic, it was suggested that bi-ethnic labor markets promote economic discrimination against minority group members, while mono-ethnic labor markets provide minorities with temporary protection from such discrimination. That is, when a minority population is able to develop independent labor markets, it is able to produce incumbents for high-status positions usually taken by members of the superordinate groups. In an ethnic enclave, the minority group has to supply workers not only for low-status, low-pay unskilled and semiskilled jobs but also to produce lawyers, doctors, nurses, mayors, and clerks.

The data analysis focused on Jews, Arabs, and their local labor markets. The findings demonstrate the significant effect of labor-market segregation on economic discrimination in Israel. Minority members who live and work in Jewish communities (residents) suffer more than any other group the detrimental consequences of both occupational and income discrimination. The vulnerability of the residents appears to be in their position in the bi-ethnic labor market. By contrast, minority members who live and work in Arab communities (segregated) are occupationally advantaged. Apparently, with the absence of competition in the ethnic enclave, they are able to attain occupational positions usually held by Jews. Furthermore, in comparison to Jews, they are able to achieve high-status positions with relatively lower education. The

commuters—those who live in Arab places but work in Jewish localities—have to settle for the low-status occupation in exchange for relatively higher income in the bi-ethnic labor market.

The findings in the paper clearly demonstrate the differential effects of local labor markets on economic discrimination. The rules by which minority members achieve socioeconomic status vary considerably across mono-ethnic and bi-ethnic labor markets. Local labor markets, thus, should be viewed not only as a result of stratification but also as one of its major causes. Indeed, they are a structural feature of the social system that exerts significant impact both on individuals' opportunities for achievement and on ethnic-linked socioeconomic inequalities.

REFERENCES

Allport, G.W. 1954. *The Nature of Prejudice*. Boston: Beacon.

Althauser, R.P. and M. Wigler. 1972. "Standardization and Component Analysis." *Sociological Methods and Research* 1:97–135.

Baron, J. and W. Bielby. 1980. "Bringing the Firms Back In: Stratification, Segmentation, and the Organization of Work." *American Sociological Review* 45:737–65.

Bibb, R. and W. Form. 1977. "The Effects of Industrial, Occupational, and Sex Stratification on Wages in Blue-collar Markets." *Social Forces* 55:474–96.

Blalock, H.M. 1967. *Toward a Theory of Minority Group Relations*. New York: Wiley.

Bonacich, E. 1972. "A Theory of Ethnic Antagonism: The Split Labor Market." *American Sociological Review* 37:547–59.

_____. 1976. "Advanced Capitalism and Black-White Race Relations in the United States: A Split Labor Market Interpretation." *American Sociological Review* 41:34–51.

Central Bureau of Statistics. 1985. *Localities and Statistical Areas: Population and Households*. Jerusalem.

Duncan, O.D. 1968. "Inheritance of Poverty or Inheritance of Race?" Pp. 85–110 in *On Understanding Poverty*, edited by Daniel P. Moynihan. New York: Basic Books.

Fischer, C. 1975. "Toward a Subcultural Theory of Urbanism." *American Journal of Sociology* 80:1319–51.

Fossett, M. 1984. "City Differences in Racial Occupational Differentiation: A Note on the Use of Odds Ratios." *Demography* 21:655–66.

Frazier, E. 1951. "The Negroes' Vested Interest in Segregation." In *Race, Prejudice and Discrimination*, edited by A. Rose. New York: Knopf.

Frisbie, W.P. and L. Neidert, 1977. "Inequality and the Relative Size of Minority Populations: A Comparative Analysis." *American Journal of Sociology* 3:1007–30.

Glenn, N. 1964. "The Relative Size of the Negro Population and Negro Occupational Status." *Social Forces* 43:42–49.

_____. 1966. "White Gains from Negro Subordination." *Social Problems* 14:159–78.

Hudson, R. and R. Kaufman 1982. "Economic Dualism: A Critical Review." *American Sociological Review* 47:727–39.

Iams, H.M. and A. Thornton. 1975. "Decomposition of Differences: A Cautionary Note." *Sociological Methods and Research* 3:341–52.

Jones, F.L. and J. Kelley. 1984. "Decomposing Differences Between Groups: A Cautionary Note on Measuring Discrimination." *Sociological Methods and Research* 12:323–43.

Kaufman, R. 1984. "A Structural Decomposition of Black-White Earning Differentials." *American Journal of Sociology* 89:585–611.

LaGory, M. and R. Magnani. 1979. "Structural Correlates of Black-White Occupational Differentiation: Will U.S. Regional Differences in Status Remain?" *Social Problems* 27:157–69.

Lewin-Epstein, N. 1985. "Neighborhoods, Local Labor Markets, and Employment Opportunities for White and Nonwhite Youth." *Social Science Quarterly* 66:163–71.

_____. 1986. "Effects of Residential Segregation and Neighborhood Opportunity Structure on the Employment of Black and White Youth." *Sociological Quarterly* 27:559–70.

Lieberson, S. 1980. *A Piece of the Pie*. Berkeley: University of California Press.

Light, I.H. 1972. Ethnic Enterprise in America: Business and Welfare among Chinese, Japanese, and Blacks. Berkeley. University of California Press.

Logan, J. 1976. "Industrialization and the Stratification of Cities in Suburban Regions." *American Journal of Sociology* 82:333–48.

_____. 1978. "Growth, Politics, and the Stratification of Places." *American Journal of Sociology* 84:404–16.

Martin, W.T. and D.L. Poston, Jr. 1972. "The Occupational Composition of White Females: Sexism, Racism and Occupational Differentiation." *Social Forces* 50:349–55.

Oaxaca, R. 1973. "Sex Discrimination in Wages." Pp. 124–51. in *Discrimination in Labor Markets*, edited by O. Ashenfelter and A. Reese. Princeton, NJ: Princeton University Press.

Parcel, T.L. 1979. "Race, Regional Labor Markets and Earnings." *American Sociological Review* 44:262–79.

Parcel, T.L. and C.W. Mueller. 1983. *Ascription and Labor Markets: Race and Sex Differences in Earnings*. New York, London: Academic Press.

Peres, Y. 1971. "Ethnic Relations in Israel." *American Journal of Sociology* 76:1029–47.

Semyonov, M. 1981. "Effects of Community on Status Attainment." *Sociological Quarterly* 22:359–72.

Semyonov, M., D.R. Hoyt, and R.I. Scott. 1984. "Place, Race and Differential Occupational Opportunities." *Demography* 21:258–70.

Semyonov, M. and A. Tyree. 1981. "Community Segregation and the Costs of Ethnic Subordination." *Social Forces* 59:649–66.

Simon, R. 1971. *Continuity and Change*. New York: Cambridge University Press.

Smooha, S. 1978. *Israel: Pluralism and Conflict*. Berkeley: University of California Press.

Spilerman, S. 1977. "Careers, Labor Market Structure and Socio-economic Achievement." *American Journal of Sociology* 83:551–93.

Spilerman, S. and J. Habib. 1976. "Development Towns in Israel: The Role of Community in Creating Ethnic Disparities in Labor Force Characteristics." *American Journal of Sociology* 81:781–812.

Spilerman, S. and R.E. Miller. 1977. "City Non-Differences Revisited." *American Sociological Review* 42:979–83.

Stolzenberg, R. 1975. "Occupations, Labor Markets and the Process of Wage Attainment." *American Sociological Review* 40:645–65.

Tienda, M. and D. Lii. 1987. "Minority Concentration and Earnings Inequality: Blacks, Hispanics, and Asians Compared." *American Journal of Sociology* 93:141–65.

Thurow, L.C. 1969. *Poverty and Discrimination.* Washington, DC: Brookings Institution.

Tyree, A. 1981. "Occupational Socioeconomic Status, Ethnicity and Sex in Israel: Consideration in Scale Construction" (in Hebrew). *Megamot* 27:7–21.

Wilcox, J. and W.C. Roof. 1978. "Percent Black and Black-White Status Inequality: Southern versus Non-Southern Patterns." *Social Science Quarterly* 59:431–34.

Williams, R.M. 1947. *The Reduction of Intergroup Tensions.* New York: Social Science Research Council.

Winsborough, H.H. and P. Dickenson. 1971. "Components of Negro-White Income Differences." Proceedings of the American Statistical Association, Social Statistics Section: 6–8.

A FLEXIBLE PROCEDURE FOR ADJUSTING RATES AND PROPORTIONS, INCLUDING STATISTICAL METHODS FOR GROUP COMPARISONS*

CLIFFORD C. CLOGG   SCOTT R. ELIASON

Pennsylvania State University

This article presents some methods based on the "purging" concept that can be used to simplify the interpretation of multivariate relationships among discrete variables. These methods can be applied in general situations where one of the variables is a specified dependent variable, another is a specified group variable, and one or more other variables are "composition" variables or covariates. The composition variables are assumed to be associated with the group levels and can thus confound the interpretation of observed group differences in the dependent variable. These purging methods can be used to adjust rates for marginal or partial two-factor interaction between the group and composition variables and for three-factor interaction among the group, composition, and dependent variables. The method of direct standardization widely used in sociology, demography, and other areas is shown to be a special case of the purging method. Statistical inference for group comparisons is facilitated by the jackknife technique. The methods are illustrated with an example dealing with the preference for additional children.

INTRODUCTION

Statistical models for analyzing multivariate relationships among discrete variables have become very popular in sociological research in the past several years. These models are generally in the log-linear family and include log-frequency models (conventional log-linear models for cell frequencies), log-odds models (logit models and more general multinominal-response models), log-odds-ratio models (association models), and log-rate models (including hazards models for discrete predictors). (See Clogg and Schockey [1988] and Clogg and Eliason [1987] for current surveys and references to the literature.) In our opinion, the advantages of these modeling techniques over simple elaboration techniques or least squares regression are clear-cut. This is certainly the case if the optimality criteria used in modern statistical science are used as a basis for comparison of methods.

On the other hand, many researchers have

been less than satisfied with the interpretive aspects of the new modeling procedures. (See Kaufman and Schervish [1986] and references cited there.) Part of the difficulty is that the natural metric of parameters in these models is a logarithmic scale. Many researchers prefer interpretations based on absolute or arithmetic scales. Interpretation can be especially difficult when polytomous variables or contingency tables of high dimension are studied. For example, when the dependent variable is polytomous, researchers ordinarily must use special contrasts of parameter values to address many of the key questions pertaining to group differences and how these differences are affected by one or more covariates. In many instances, a thorough evaluation of group differences is simply avoided because of the complexities involved. Some researchers rely instead on goodness-of-fit statistics or nested comparisons of chi-squared values to examine the relative contribution of various parameters. While this strategy has its merits, it does not lead to concise descriptions of findings: it is like summarizing regression results using only multiple and partial R-squared values and ignoring the estimated values of the regression coefficients.

A much older technique for analyzing group differences in rates and proportions is the method of standardization; its primary benefit is ease of interpretation. Standardized

rates can be compared to each other and to "crude rates" in a variety of ways, all of which rely on the intuitively appealing metric of a rate, a proportion, or a percentage distribution. This method was introduced to sociologists in Rosenberg (1962) as "test-factor standardization"; Davis (1984) extended this method in several important ways. Little and Pullum (1979) considered the relationship between standardization and the log-linear model (see also Smith [1977]). Stinchcombe (1968) used standardization explicitly in his treatment of demographic explanation of social facts. Standardization has a long history in demography. Kitagawa (1955, 1964), Keyfitz (1968), Shryock and Siegel (1980), Spiegelman (1968), and Wunsch and Termote (1978) summarize standardization methods for demographic applications. The latter source actually features the logic of standardization as the central characteristic of demographic methodology.

Standardization methods play a central role in other areas of statistics; Fleiss (1981) and Freeman and Holford (1980) are valuable guides to the literature. In biostatistics, the same technique has been used frequently as a method for bias reduction. For example, standardization methods are compared to matching procedures and to analysis of covariance in Anderson (1980). Bishop, Fienberg, and Holland (1975) considered standardization methods in their fundamental treatise on the log-linear model. Their initial work on the log-linear model, carried out in connection with the National Halothane Study of the 1960s, was actually motivated by the need to create "smoothed" cell frequencies so that standardization methods could be applied to evaluate the importance of several risk factors. Haberman (1979, ch. 9) deals with modifications of the Deming method of adjustment to marginal distributions that are also closely related to conventional standardization techniques. The emphasis placed on standardization methods of various kinds in sociology, demography, and other areas is undoubtedly associated with the benefits of standardization as a descriptive method. Interest in standardization techniques has not abated with the development of formal modeling procedures for the analysis of categorical data.

Our purpose is to integrate the modeling capabilities associated with log-linear analysis with the descriptive features of standardization methods. The *purging* technique given in Clogg (1978, 1979) and Clogg and Schockey (1985) provides the foundation for the methods that are introduced. We show how many of the more conventional procedures for standardization can be viewed as a special case of general procedures based on the purging concept. Statistical methods for group comparisons are considered as well. These are based on the jackknife method of assessing precision of adjusted rates and various kinds of contrasts among them. Before presenting the purging method and generalizations of it, we consider an example that illustrates some of the advantages of our general approach.

AN EXAMPLE

Table 1, from Rindfuss and Bumpass (1978, p. 49), presents data on the relationships between age and parity (number of children) and the desire for more children. *Age* levels are indexed by $i = 1, \ldots, 5$ corresponding to age groups 20–24, . . ., 40–44, respectively; *parity* levels are indexed by $j = 1, \ldots, 4$ corresponding to parity 1, 2, 3, and 4 or more, respectively; and the levels of the dichotomous *desire* variable are indexed by $k = 1$ (Yes) and $k = 2$ (No). The age by parity by desire ($5 \times 4 \times 2$) table would ordinarily be viewed in an asymmetric way, with age and parity as specified explanatory variables and desire as the specified dependent variable. Most analysts would be interested in the effects of age (adjusted for parity composition) and the effects of parity (adjusted for age composition); the possibility of interaction effects (between age and parity) and their role in "specifying" age and parity effects would also be examined. Various models for a

Table 1. Desire for More Children (Yes or No) by Maternal Age and Parity, for Women with at Least One Child in 1970

	Parity							
	1		2		3		4+	
Age	Yes	No	Yes	No	Yes	No	Yes	No
20–24	327	36	133	138	22	46	10	17
25–29	160	48	118	228	58	156	29	123
30–34	54	42	43	219	29	165	34	190
35–39	12	47	14	158	12	156	12	227
40–44	5	43	4	142	7	128	13	198

Source: Rindfuss and Bumpass (1978, p. 49), tabulated from the 1970 National Fertility Survey.

dichotomous dependent variable would thus be considered to explain the dependence of desire on age and parity.

Let f_{ijk} denote the observed frequency in cell (i,j,k) in Table 1, and let F_{ijk} denote the expected frequency under some model. Let $f_{ij+} = f_{ij1} + f_{ij2} = N_{ij}$ denote the total number of observations in age i and parity j. Most models that would be used in this case condition on the observed values of the N_{ij}, implying that $N_{ij} = F_{ij+} = \hat{F}_{ij+} = \hat{F}_{ij1} + \hat{F}_{ij2}$, where circumflexes denote sample estimates. For example, from Table 1 we find that $f_{111} = 327$ and $f_{112} = 36$, giving $N_{11} = 327 + 36 = 363$. The observed proportion in age level 1 and parity 1 desiring more children ($k=1$) is $p_{11} = 327/363 = .901$; the observed odds is $p_{11}/(1 - p_{11}) = 327/36 = 9.08$, and the observed log-odds or logit is $\log(9.08) = 2.21$. These three quantities (proportions, odds, and logits) are predicted in various ways by formal models for the contingency table.

Three Different Models

One popular model generally associated with Grizzle, Starmer, and Koch (1969) is the linear-probability model. It is formally equivalent to a linear-regression model with a dichotomous (dummy) dependent variable. Let $P_{ij} = F_{ij1}/N_{ij}$ denote the expected probability that level k (of desire) $= 1$ when age and parity are at levels i and j, respectively. The saturated model is

$$P_{ij} = \alpha + \beta_i + \gamma_j + \delta_{ij}, \qquad (1)$$

where α is a constant, β_i and γ_j denote main effects of age and parity, and δ_{ij} denotes interaction or departure from additivity. Because $P_{ij} = E(f_{ij1}/N_{ij})$ and $V(f_{ij1}/N_{ij}) = P_{ij}(1 - P_{ij})/N_{ij}$, some correction for heteroskedasticity is required. The model in equation (1) can be .estimated by weighted-least-squares methods with widely available software.

The logit model of Bishop (1969) or Goodman (1978) is

$$\log[P_{ij}/(1 - P_{ij})] = \alpha + \beta_i + \gamma_j + \delta_{ij}, \quad (2)$$

where $\log[P_{ij}/(1 - P_{ij})] = \log(F_{ij1}/F_{ij2})$ is the log-odds that $k = 1$ when age and parity are at levels i and j, respectively. The parameters

on the right-hand side of equation (2) have the same interpretation as previously, although they pertain to effects on logits rather than effects on probabilities. The parameters in equation (2) are not equal to the parameters in equation (1), except in some trivial cases. This model can be estimated in a variety of ways, directly with a program for logit analysis or indirectly with a program for log-linear models for contingency tables. Maximum likelihood estimates are available with most software packages.

Often, the proportion P_{ij} is referred to as a *rate*, in which case the model in equation (1) is a linear model for rates. A simple alternative to both equation (1) and equation (2) is the *log-rate model*, which corresponds to an assumption that the explanatory variables affect the P_{ij} in a multiplicative instead of an additive way. This model will be written as

$$\log(F_{ij1}/N_{ij}) = \alpha + \beta_i + \gamma_j + \delta_{ij}. \quad (3)$$

Haberman (1978, ch. 1) considers a model similar to this one. Note that only the expected frequency for level $k = 1$ (and age i and parity j) is included in the left-hand side. The totals N_{ij} appear as fixed *weights* that adjust this frequency and, in fact, convert it into a rate. (More generally, F_{ij1} can represent an event count with N_{ij} denoting the fixed exposure under which those events occurred.) The model in equation (3) is a special kind of weighted log-linear model, and it can also be estimated easily with existing software. (For example, in SPSS[x] LOGLINEAR, the N_{ij} are regarded as cell weights in the CWEIGHT option; in GLIM the N_{ij} would be used in an "offset" command.) See Clogg and Eliason (1987) for discussion and Allison (1984) for some closely related results. Importantly, the totals N_{ij} are regarded as fixed (or are "fitted") with each model. This assumption is explicit with the model in equation (3), where these totals become fixed weights to which each "event count" is adjusted (to produce a rate). To facilitate later developments, we note that the N_{ij} merely represent the cross-classification of the two explanatory variables, or the *marginal* two-way contingency table obtained by collapsing over the levels of the dependent variable, $N_{ij} = f_{ij+}$.

Clearly, other models could be entertained for the data in Table 1, but we will consider

just these three to make the point: formal statistical modeling does not lead automatically to a clear-cut *description* of the data.

Modeling the Data: Estimating the Three Models

Table 2 gives parameter values and tests for interaction effects when the three models above are applied to the data in Table 1. The chi-squared values reported at the bottom of the table all pertain to the null hypothesis that $\delta_{ij} = 0$ or all i and j, or that age and parity effects are additive (on the scale appropriate for each model). The linear probability model of equation (1) was estimated with weighted least squares; the chi-squared value (Wald statistic) for interaction is 111.67 on 12 degrees of freedom, so this model cannot be replaced by a model with main effects. The logit model of equation (2) also has significant interaction effects; testing the model of main effects (setting the interaction terms to zero) yields a likelihood-ratio chi-squared value of 30.82 on 12 degrees of freedom,

which is still highly significant. The log-rate model of equation (3) can actually be replaced by the model with main effects: the chi-squared value is 15.87 (p − value = .20) for the main-effects model.

Consider first the inferences from, the log-rate model, which are the easiest to describe because the parity × age interaction is not required. Suppose that we were interested in studying the differences among parity groups adjusted for age effects. The γ_j parameters in the model represent the effects of parity adjusted for age effects; the estimated vales of these parameters are .740, −.041, −.289, and −.410 [= −(.740 − .041 − .289)] for parities 1, 2, 3, and 4+. The contrast between parity 1 and parity 4+, for example, is thus .740 − (−.410) = 1.150; exp(1.150) = 3.158 is the estimated "relative risk ratio" adjusted for age effects. Further manipulations of the parity effects of this kind can be used to describe the parity differences adjusted for age effects, but it would require considerably more work to

Table 2. Linear-Probability, Logit, and Log-Rate Models Applied to the Data in Table 1

	Model			
Parameter Value[a]	Saturated Linear Probability	Saturated Logit	Saturated Log-Rate	Main-Effects Log-Rate
α	.267	−1.323	−1.737	−1.724
β_1	.255	1.549	1.003	.929
β_2	.126	.851	.662	.688
β_3	−.010	.114	.195	.184
β_4	−.165	−1.000	−.696	−.715
γ_1	.241	1.352	.778	.740
γ_2	−.046	−.340	−.203	−.041
γ_3	−.093	−.464	−.249	−.289
δ_{11}	.138	.629	−.148	—[c]
δ_{12}	.015	.078	.226	
δ_{13}	−.104	−.499	−.145	
δ_{21}	.135	.324	.035	
δ_{22}	−.006	.153	.202	
δ_{23}	−.029	−.053	.020	
δ_{31}	.064	.108	.189	
δ_{32}	−.047	−.079	−.062	
δ_{33}	−.014	−.066	−.109	
δ_{41}	−.139	−.394	−.062	
δ_{42}	.026	.240	.127	
δ_{43}	.063	.222	.043	
Chi-square[b] for interaction	111.67	30.82	15.87	15.87

[a] Under each model, parameter values are identified by standard restrictions, $\Sigma_i \beta_i = \Sigma_j \gamma_j = \Sigma_i \delta_{ij} = \Sigma_j \delta_{ij} = 0$. Only identified values are reported; the others can be determined from the ones reported. For example, $\hat{\beta}_5 = -(\hat{\beta}_1 + \ldots + \hat{\beta}_4)$, $\hat{\gamma}_4 = -(\hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3)$, $\hat{\delta}_{14} = -(\hat{\delta}_{11} + \hat{\delta}_{12} + \hat{\delta}_{13})$, etc.

[b] Wald statistic for interaction is used for the linear-probability model; likelihood-ratio statistics from the main-effects model are used for the other two models. Both statistics have approximate chi-squared distributions under the paricular null model of no interaction. There are 12 degrees of freedom for interaction (equal to the number of nonredundant δ_{ij} values).

[c] Interactions restricted to zero under the main-effects log-rate model.

convert the parity effects into interval esti-
mates for the rates or to test specific
hypotheses about parity effects. In this
particular case, the fact that the additive
model ($\delta_{ij} = 0$ for all i and j) provided an
acceptable level of fit is a definite advantage.
It would be more difficult to assess parity
effects if there were interaction. Another
difficulty is that this example pertains only to
a dichotomous dependent variable. While
both the linear-probability model and logit
model can be generalized easily to deal with
polytomous dependent variables, it would be
more difficult to generalize the log-rate model
in this way. We are not aware of such
generalizations in the literature.[1]

With both the linear-probability model and
the logit model, the evidence for interaction
effects is clear-cut. Had we used the logit
model, for example, not much further simpli-
fication of the data would be possible, and the
researcher would have to explain the age and
parity effects taking account of the interac-
tions. There are many methods available for
coming to grips with the interaction effects,
including graphical methods (Fox 1987).
There is no question that experienced analysts
could make sense of the parameter values in
Table 2 after some work. Rather than go into
details, we consider instead the alternative
strategy of summarizing the relationship in
terms of adjusted rates.

*Describing the Data: Interpretations Based
on Adjusted Rates*

Crude and adjusted rates (or percentages)
appear in Table 3. First, consider the
differences among parity levels in the percent-
age desiring more children. The parity levels
are here regarded as groups (group variable G
= parity) and age levels are regarded as a
composition variable (composition variable C
= age). The "crude" percentages can be

---

[1] The log-rate model provides a natural way to simul-
taneously standardize and model rates in cases where the
dependent variable is dichotomous. The quantity F_{ij1}/N_{ij}
is formally equivalent to the composition-specific rate $r_{i(k1)}$
defined in equation (4). We can model directly standard-
ized rates (also defined below) given a set of standard
weights c_j^* by using $\log[F_{ij1}/(N_{ij}/c_j^*)] = \log[F_{ij1}c_j^*/N_{ij}]$ on
the left-hand side of equation (3). While this approach is
attractive in certain respects, the approach of the present
paper appears to be both more general and more flexible.
For example, this particular modification of the log-rate
model is not directly applicable to the analysis of polyto-
mous dependent variables.

Table 3. Crude and Adjusted Percentage Desiring More
Children by Parity Group

Parity Group	Crude Percentage[a]	Adjusted Percentage[b]
1	72.1%	57.7%
2	26.1	20.1
3	16.4	18.2
4+	11.5	16.9
Contrasts:[c]		
1 vs. 2	46.0	37.6
	(2.1)[d]	(4.1)
2 vs. 3	9.6	1.9
	(1.8)	(2.9)
3 vs. 4+	4.9	1.2
	(1.7)	(2.6)
1 vs. 4+	60.6	40.8
	(1.9)	(3.8)

[a] Observed percentage desiring more children at each
parity level (from the two-way cross-classification of
parity and desire).
[b] Adjusted percentages obtained from the method of
"Partial CG and CGD" purging described in the text.
[c] Simple difference in percentages, parity 1 minus
parity 2, etc.
[d] Standard errors (in parentheses) obtained from the
jacknife method described in equations (24) and (25).

obtained from the two-way cross-classification
of parity and desire: 72.1 percent of females
at parity 1 desire more children; this declines
to 26.1 percent, 16.4 percent, and 11.5
percent for parities 2, 3, and 4+. A useful
way to summarize the observed parity differ-
ences is to contrast each adjacent pair of
groups (1 vs. 2, 2 vs. 3, 3 vs. 4+). The
overall difference between parity group 1 and
parity group 4+ is also reported. The values
of these contrasts (simple arithmetic differ-
ences in percentages) along with their stan-
dard errors appear in the lower panel of Table
3. For example, the difference between parity
3 and parity 4+ is 4.9 percent with a standard
error of 1.7 percent: the observed difference
divided by its standard error is about 2.8, so
this difference is statistically significant. The
difference between parity 1 and parity 2 is
dramatic (46.0 percent with a standard error
of 2.1 percent, but all differences in parity
groups are highly significant when the crude
percentages are considered.

Using one of the rate-adjustment methods
presented below (the "Partial CG and CGD"
method), the adjusted rates and contrasts
among them in the second column of Table 3
were obtained. These rates control for the
confounding effect of age composition, which
is necessary because age and parity are
(positively) associated. They also adjust for
three-factor interaction among the three vari-

ables, which is equivalent to two-factor inter-action in the logit model of equation (2). (Procedures for determining whether it is sensible to adjust for three-factor interaction appear below; in this case it is appropriate to summarize group differences with this method.) The adjusted rates reveal parity differences that are still substantial, but they have been reduced a great deal by controlling for age composition. Note that the differences between parity 2 and parity 3 and between parity 3 and parity 4+ are now less than 2 percent. These estimated differences are actually less than their corresponding standard errors. The difference between parity 1 and parity 2 is now 37.6 percent (compared to the crude difference of 46.0 percent); adjustment for age effects reduces this difference by about 8 percent in absolute terms. In other words, the only salient parity difference in the desire for more children is that between parities 1 and 2 once age composition is controlled.

We next consider the age differences in the desire for more children. Table 4 gives the crude and adjusted percentages corresponding to those in Table 3. Now age is taken as the group variable and parity is taken as the composition variable. The adjustment process here considers age differences adjusted for parity composition. (It is useful to note that the designation of the group and composition variables is arbitrary; the two types of variables can be interchanged.) The crude percentages show a monotonic decline in the desire for more children with increasing age. While the adjusted percentages also exhibit a monotonic decline with increasing age, the

age differences are reduced after controlling for parity composition. For example, the adjusted difference between age groups 20–24 and 25–29 is 17.2 percent; the crude difference is 27.8 percent, so adjusting for parity composition reduces this difference by over 10 percent in absolute terms. The difference between age groups 25–29 and 30–34 is reduced by about 4 percent (from 19.1 percent to 15.2 percent). The differences among the oldest three age groups are not affected very much by adjusting for parity composition; in fact, the differences are slightly larger for the adjusted percentages.

Compare the results in Table 3 and 4, based on adjusted percentages, with the results in Table 2, based on the parameter values in formal statistical models. In our opinion, interpreting the data is much simpler with the results based on adjusted rates. Of course, after looking at the adjusted rates and various contrasts among them, it is possible to see patterns in the estimated parameter values in Table 2, but it would be very difficult to obtain the simplified description of parity and age effects on the desire for more children that is provided by the adjusted rates. It has been our experience that adjusted rates of this kind usually lead to simpler interpretations of multivariate relationships among discrete variables than do the corresponding model parameters. In the next section we describe the rationale for the rate-adjustment procedures used to create Tables 3 and 4.

A GENERAL FRAMEWORK FOR ADJUSTING RATES

Notation and Definitions

Let C, G, and D denote the *composition*, *group*, and *dependent* variables, each of which is discrete. Let I, J, and K denote the number of levels of C, G, and D, respectively. As in the example discussed above, C and G might be interchangeable. Each variable could represent crossed or nested combinations of other variables; for example, C could refer to age-by-sex. In fact, the procedures given here are fully multivariate (not just suited for three specific variables), a point to which we return later. Note that the number of levels of each variable can be greater than or equal to two; we do not restrict attention to just dichotomous variables or to just situations with dichotomous dependent

Table 4. Crude and Adjusted Percentage Desiring More Children by Age Group

Age Group	Crude Percentage	Adjusted Percentage
1. 20–24	67.5%	54.5%
2. 25–29	39.7	37.4
3. 30–34	20.6	22.2
4. 35–39	7.8	8.6
5. 40–44	5.4	5.3
Contrasts:		
1 vs. 2	27.8	17.2
	(2.4)	(3.8)
2 vs. 3	19.1	15.2
	(2.2)	(2.5)
3 vs. 4	12.8	13.7
	(1.8)	(2.1)
4 vs. 5	2.5	3.3
	(1.4)	(1.7)

See notes to Table 3.

variables. The cross-classification of C, G, and D gives an $I \times J \times K$ contingency table. Many research questions in sociology and other areas can be expressed in terms of contingency tables of this kind, i.e., cross-classifications of specified composition, group, and dependent variables. The main question asked in such situations is this: to what extent do the groups differ on levels of the dependent variable when compositional differences among the groups are controlled? We seek a general strategy that makes the controlling procedure explicit, that recognizes sampling variability in the data, and that produces summary group rates consistent with the model specified for the data and the substantive question to be examined.

Let f_{ijk} denote the observed frequency in cell (i,j,k) of the three-way contingency table, and let F_{ijk} denote the corresponding expected frequency under some model, as before. We let $\hat{F}_{ijk}$ denote the maximum likelihood estimate of F_{ijk} under the model. The unrestricted (or saturated) model gives $F_{ijk} = f_{ijk}$. If the $\hat{F}_{ijk}$ are obtained from a restricted (or unsaturated) model, we say that the data have been *smoothed*. In this case, the discrepancies between the observed frequencies f_{ijk} and the estimated expected frequencies $\hat{F}_{ijk}$ would ordinarily be attributed to sampling error.

Composition-specific rates are defined as

$$r_{ij(k)} = F_{ijk}/F_{ij+},\qquad (4)$$

where "+" denotes summation over the given subscript. (Note that $F_{ij+} = N_{ij}$ is the total in the ith composition category and the jth group.) From equation (4) we see that $r_{ij(k)}$ is the rate of prevalence of the kth category of D in the ith category of C and the jth group. The *crude rate* (or unadjusted rate) of the kth category of D in the jth group is

$$r_{.j(k)} = F_{+jk}/F_{+j+},\qquad (5)$$

where F_{+j+} is the total number of observations in the jth group. Note that crude rates are obtained from the *marginal* two-way table cross-classifying G and D (F_{+jk}).

Standardized Rates: Direct Method

Using equation (4), equation (5) can be rewritten as

$$r_{.j(k)} = \Sigma_i c_{ij} r_{ij(k)}\qquad (6)$$

where

$$c_{ij} = F_{ij+}/F_{+j+}\qquad (7)$$

is the fraction of the jth group in the ith composition category, with $\Sigma_i c_{ij} = 1$ for each j. Equation (6) shows that the crude rates are weighted averages of the $r_{ij(k)}$ where the weights c_{ij} can vary across groups for some or all levels i of C. Note that the weights are obtained from the marginal two-way cross-classification of C and G with cell entries F_{ij+}.

In view of equation (6), the crude rates depend on two factors: the distribution of C in the jth group and the composition-specific rates. The method of standardization controls for compositional differences among the groups by defining a *standard set of compositional weights* c_i^s, with $\Sigma c_i^s = 1$. Standardized rates (direct method) based on the c_i^s are defined as

$$r_{.j(k)} = \Sigma_i c_i^s r_{ij(k)}.\qquad (8)$$

Group differences in the composition are "controlled" because the same set of compositional weights are applied to each group (Kitagawa 1964). The standardized rates depend on the choice of the c_i^s; much of the literature on standardization pertains to problems in selecting an "optimal" set of weights. Popular choices for the c_i^s include

a. $c_i^s = c_{ij'}$ (the j'th group serves as the "standard")
b. $c_i^s = \Sigma_j c_{ij}/J$ (arithmetic average of the c_{ij} across groups)
c. c_i^s derived from some external standard, such as a stable age distribution

Choice a. is relevant when a particular group is a natural reference category to which other groups should be compared. Choice b. leads to the components methods of Kitagawa (1955) when only two groups ($J = 2$) are compared (also see Das Gupta [1978]). Choice c. is often preferred in demographic studies on fertility or mortality. Under either of the first two choices, the c_i^s are determined from the data and thus will be random rather than fixed. This issue complicates the statistical theory appropriate for standardized rates; see Keyfitz (1966a, 1966b, 1968) and Freeman and Holford (1980). Many researchers have noted that problems in selecting a

standard set of weights arise in cases where there is three-factor (C-G-D) interaction in the data. This problem will be addressed below.

Two Models Relevant for the Rate-Adjustment Problems

We indicated above that composition-specific rates are based on the frequencies F_{ijk} in the three-way table cross-classifying C, G, and D and that the composition weights c_{ij} on which the crude rates depend are based on the frequencies F_{ij+} in the C by G (two-way) marginal table. Modeling either or both of these sets of frequencies is the first step in the proposed framework.

A general multiplicative (or log-linear) model for the three-way table (Goodman 1978) is

$$F_{ijk} = \tau \, \tau_i^C \tau_j^G \tau_k^D \, \tau_{ij}^{CG} \tau_{ik}^{CD} \tau_{jk}^{GD} \tau_{ijk}^{CGD}. \quad (9)$$

Clogg (1978) discusses the meaning of the parameters in the same situation considered here, so a brief summary will be sufficient. The three-factor interaction is represented by the set of IJK τ_{ijk}^{CGD} parameters. The case where $\tau_{ijk}^{CGD} \neq 1$ for at least some (i,j,k) can be interpreted as saying that group differences in the levels of the dependent variable differ across levels of the composition variable. In cases where the dependent variable is dichotomous, we can form the odds F_{ij1}/F_{ij2} and transform the tau parameters into parameters for the odds. The δ_{ij} values for interaction in the logit model of equation (2) are just 2 $[\log(\tau_{ij1}^{CGD})]$. In fact, all parameters in equation (9) involving variable D relate directly to the parameters of the logit model; see Fienberg (1980) for details. When the τ^{CGD} values are null (equal to 1), the two-factor tau parameters measure the partial association between the two given variables "controlling" for levels of the third variable. Thus, the τ^{CG} measure the partial association between C and G, the τ^{GD} measure the partial association between G and D, and the τ^{CD} measure the partial association between C and D. These three two-factor taus refer to composition-group interaction (in the three-way table), to group differences in D (adjusted for C), and to compositional effects on D (adjusted for G). Each of these interactions has an effect on the crude rates, as can be seen by substituting equation (9)

into equation (5). The crude rates thus depend on the partial composition effects and the partial group effects, but they also depend on the partial composition-group interaction. The τ^{CG} confound the interpretation of crude rates, an observation that leads to the basic idea of purging given below.

A general model for the two-way C-by-G table is

$$F_{ij+} = \gamma \, \gamma_i^C \, \gamma_j^G \, \gamma_{ij}^{CG}. \quad (10)$$

This is just a saturated multiplicative model for the two-way table; the γ_{ij}^{CG} refer to interaction (lack of independence) between C and G. The γ_{ij}^{CG} of equation (10) will be called *marginal CG* interactions, while the τ_{ij}^{CG} of equation (9) will be called *partial CG* interaction, using obvious nomenclature. Note that the γ_{ij}^{CG} values will be different from the τ_{ij}^{CG} values, except in some special cases (Bishop, Fienberg, and Holland 1975; Whittemore 1978).

Analysis of restricted (or unsaturated) models related to the ones above can play a role in assessing group differences in the levels of the dependent variable and in determining how or whether to calculate adjusted rates. The following three models are directly relevant:

H_1: No three-factor interaction ($\tau^{CGD} = 1$).
H_2: Conditional independence between G and D given the levels of C ($\tau^{GD} = \tau^{CGD} = 1$)
H_3: Independence between C and G ($\gamma^{CG} = 1$).

Model H_1 can be viewed as a test for the standard problem referred to earlier. If H_1 is true, the G-D partial association is constant across levels of C, in which case different sets of standard weights c_i^j will at least preserve the relative ordering of the groups as assessed from the adjusted rates. Model H_2 is the null model that should be used to test for group differences in levels of D. If this hypothesis is true, *any* standardization or adjustment procedure will lead to across-group homogeneity in the adjusted rates. Model H_3 is relevant because it tests whether the c_{ij} are homogeneous across groups, in which case standardization is not required.

These models are relevant for other purposes as well. For example, if model H_1 cannot be rejected, it would be preferable to use the fitted frequencies $\hat{F}_{ijk}$ under H_1 as a basis for inference, including the calculation

of adjusted rates. There are two main reasons for doing this: (1) the $\hat{F}_{ijk}$ and quantities based on them (such as adjusted rates) will have smaller sampling variability than the observed frequencies, if H_1 is true (Altham 1984); (2) model H_1 smooths the data of irregularities and usually removes zero frequencies that cause special difficulties. Many other restricted models can be used to describe the data, including models that take account of possible ordering of categories (Goodman 1984). These other models merely restrict the parameter values in equations (9) or (10) further. In the following, we assume that some particular restricted or unrestricted model has already been found for each contingency table. The idea is that the data are modeled in whatever way is appropriate for the problem at hand using the general multiplicative models in equations (9) and (10). The description of the data offered by the rate-adjustment procedures introduced next can be viewed as a logically distinct set of operations. In fact, the adjusted rates are functions of the parameter values in either equations (9) or (10), or both.

The Purging Method

Using either or both models above is valuable because it focuses attention on possible interactions that can confound the interpretation of group differences. The purging method has the following steps. First, the confounding interactions are identified; usually the researcher will wish to consider several possible such interactions. The $\tau_i^{C}{}_j^{G}$, the τ_{ijk}^{CGD}, and the $\gamma_{ij}^{C}{}^{G}$, each of which refers to interaction between composition and group, must be taken into account. Note that this step assumes that a model, possibly incorporating restrictions on these or other parameter values, has already been selected. Second, *purged frequencies* F^*_{ijk} are calculated by removing the effects of the confounding interactions. Third, adjusted rates are calculated from the purged frequencies using the formula

$$r^*_{.j(k)} = F^*_{+jk}/F^*_{+j+} , \qquad (11)$$

which is just the formula for crude rates applied to the purged frequencies; see equation (5). The adjusted rates are then interpreted in various ways, for example, by comparing them to the crude rates to

determine the effects of the confounding interaction on the crude rates. Note that the emphasis in the above procedure is different from that of conventional standardization in important ways. The procedure requires an explicit model for the data (for one or both tables considered), and the process is based on the expected frequencies (which, in turn, depend on the parameter values) under that model. The focus is on creating purged frequencies and adjusted rates, not on decomposing differences in rates (so-called components methods). Methods that emphasize decomposition (Kitagawa 1955; Das Gupta 1978; Burch and Madan 1986) are difficult to justify in sampling situations, in cases where there are more than two groups, and in cases where relative comparisons of rates (such as ratios of rates) rather than arithmetic differences are of interest. In contrast, the purging method is suitable for sampling situations, for multiple groups, for polytomous dependent variables, and for cases where both relative and absolute comparisons among groups are of interest.

Partial CG Adjustment. Clogg's (1978, 1979) method uses the model-in equation (9) for the three-way table, takes the $\tau_i^{C}{}_j^{G}$ parameters as the set of confounding interactions, and defines purged frequencies as

$$F^*_{ijk} = F_{ijk}/\tau_i^{C}{}_j^{G}. \qquad (12)$$

Because the τ^{CG} values measure the partial association between the composition and group variables, these purged frequencies reflect adjustment for "partial CG" effects. Adjusted rates calculated from equation (11) depend only on the values of the other parameters in the model. See Keppel (1981) and Clogg and Shockey (1985) for applications of partial CG purging. See Shah (1986) for discussion and comparison with standardization. Under the special case where $\tau_i^{C}{}_j^{G}{}_k^{D} = 1$ for all i, j, and k (no three-factor interaction), the F^*_{ijk} of equation (12) satisfy the following condition: C and G are independent of each other for each level of D. In other words, the purging procedure of equation (12) in this case produces cell frequencies where the distribution of C is the same for each group within categories of the dependent variable. This can be contrasted with the method of marginal-CG purging described below.

Partial CG and CGD Adjustment. This

method makes a simultaneous adjustment for both partial CG and three-factor CGD interaction. That is, it identifies both τ^{CG} and τ^{CGD} as confounding interactions. Purged frequencies are here given by

$$F^*_{ijk} = F_{ijk}/(\tau^{CG}_{ij}\tau^{CGD}_{ijk}).\qquad(13)$$

The purged frequencies in equation (13) depend only on the CD and GD interactions in the model of equation (5), the latter of which describe group differences in levels of D. The F^*_{ijk} here satisfy the condition that composition levels and group levels are independent for each level of the dependent variable. Adjusted rates calculated from equations (12) and (13) can be compared to each other to determine whether three-factor interaction has a substantively important influence on the description of group differences. If the two sets of adjusted rates are in essential agreement, then three-factor interaction can be ignored as a confounding factor. Differences between the two sets of adjusted rates can be used to isolate the effect of three-factor interaction and simplify the interpretation. (This will be illustrated below with the data in Table 1.)

Often the existence of three-factor interaction is viewed as a condition where group differences cannot be summarized in terms of adjusted rates. A key reference is Kitagawa (1966); also see Bishop, Fienberg, and Holland (1975), Spiegelman (1968), and Fleiss (1981). Of course, when three-factor interaction exists, the group differences in the levels of D depend on the level i of C, so some information will be lost if summary rates are calculated, regardless of the adjustment procedure used. But by comparing adjusted rates calculated from both of the above methods, it is possible to determine how much information is lost and to determine whether the three-factor interaction is so perverse that summary rates will be grossly misleading. In some cases, the researcher will find that the two sets of rates differ appreciably; this might suggest possible ways to stratify composition levels for separate analyses. But in other cases (e.g., with Table 1), it will be possible to effectively rule out the three-factor interaction as a confounding factor. This use of the partial-CG-and-CGD method of purging illustrates an important point. It is not desirable to adhere to just one method of rate adjustment—results from two

or more methods should be compared to each other as a kind of cross-validation exercise.

Marginal CG Adjustment. An alternative approach is to focus on the interactions between C and G in the two-way cross-classification of these variables. In other words, the γ^{CG}_{ij} of equation (10) are regarded as the confounding interactions. Purging the data of this interaction is equivalent to adjusting for across-group heterogeneity in the distribution of the composition variable. The purged frequencies are

$$F^*_{ijk} = F_{ijk}/\gamma^{CG}_{ij}.\qquad(14)$$

Using equations (10) and (14) we find that $F^*_{ij+} = F_{ij+}/\gamma^{CG}_{ij} = \gamma\,\gamma^C_i\,\gamma^G_j$, which shows that the marginal $C \times G$ table calculated from the purged frequencies satisfies the relationship of row-column independence. The purging in equation (14) creates a set of frequencies where there is no association between the two explanatory variables C and G. Using an analogy to regression, the explanatory variables are not correlated in the set of purged frequencies. In other words, the purged frequencies obtained from this method satisfy the conditions that would be obtained in a classical experiment where orthogonality of explanatory variables is built in by design. In contrast, the previous method (equivalent to partial CG purging if there is no three-factor interaction) creates purged frequencies where the two explanatory variables are not associated at *each level* of the dependent variable, in which case variable C can be collapsed. The choice between marginal-CG purging and partial-CG purging should depend on whether an inference based on orthogonalized explanatory variables or an inference based on a collapsible three-way table is sought. See Clogg, Shockey, and Eliason (1987) for further details.

Marginal CG and CGD Adjustment. There are several ways to purge frequencies for both marginal CG interaction and three-factor interaction. The most natural is

$$F^*_{ijk} = F_{ijk}/(\gamma^{CG}_{ij}\tau^{CGD}_{ijk}).\qquad(15)$$

Clearly, the F^*_{ijk} do not exhibit three-factor interaction. However, the F^*_{ij+} need not satisfy the condition of row-column independence. The purging process in equation (15) must be understood in terms of a simultaneous adjustment for the γ^{CG}_{ij} and the

$\tau_i^C{}_j^G{}_k^P$ values. There are other possibilities (see Clogg, Shockey, and Eliason 1987). An important use of this method is to compare results obtained from it with those obtained from the marginal-CG method to determine the effect of three-factor interaction, just as the partial-CG method can be compared with the partial-CG-and-CGD method.

Purging and Rate Adjustment Using a Standard Group

Often a particular group is a natural reference category to which other groups should be compared. Suppose that the sth group is chosen as the standard. This ordinarily mean that (a) crude rates and adjusted rates will be identical for the sth group; (b) confounding interactions to be controlled will be defined with the sth group as a reference category; and (c) group comparisons will be made by contrasting each group's rates to those of the standard group. Each of the above purging methods can be modified to incorporate a standard group. For example, partial-CG purging with a standard group replaces equation (12) with

$$F_{ijk}^* = F_{ijk}\,\tau_i^C{}_s^G/\tau_i^C{}_j^G. \qquad (16)$$

Using equation (9) we find that $F_{isk}^* = F_{isk}$, or purged frequencies are identical to the original frequencies in the sth (standard) group. This implies that crude and adjusted rates will be identical for the standard group. From equation (16) we see that $\tau_i^C{}_{j*}^G = \tau_i^C{}_j^G/\tau_i^C{}_s^G$ is the set of interaction values that is purged, which are defined as a contrast between the jth and the sth group; condition (b) is thus satisfied. Adjustment factors for the other purging methods, when a standard group is used, are

$\tau_i^C{}_s^G\ \tau_i^C{}_{s}^G{}^P\ /(\tau_i^C{}_j^G\ \tau_i^C{}_j^G{}^P)$ [method of partial CG and CGD purging]

$\gamma_i^C{}_s^G/\gamma_i^C{}_j^G$ [method of marginal CG purging]

$\gamma_i^C{}_s^G\,\tau_i^C{}_s^G{}^P/(\gamma_i^C{}_j^G\,\tau_i^C{}_j^G{}^P)$ [method of marginal CG and CGD purging]

Marginal CG Purging and Direct Standardization

We next show that marginal CG purging is equivalent to the method of direct standardiza-

tion. Purged frequencies with this method are determined from equation (14). First note that composition-specific rates (see equation [4]) are not affected by the purging process, i.e.,

$$
\begin{aligned}
r_{ij(k)}^* &= F_{ijk}^*/F_{ij+}^* \\
&= (F_{ijk}/\gamma_i^C{}_j^G)/F_{ij+}/\gamma_i^C{}_j^G \\
&= r_{ij(k)}. \qquad (17)
\end{aligned}
$$

Next note that the adjusted rates calculated from equation (14) can be expressed as

$$
\begin{aligned}
r_{j(k)}^* &= F_{+jk}^*/F_{+j+}^* \\
&= \Sigma_i\,(F_{ij+}^*/F_{+j+}^*)\,(F_{ijk}^*/F_{ij+}^*). \\
&= \Sigma_i\,c_{ij}^*\,r_{ij(k)}. \qquad (18)
\end{aligned}
$$

The last equality follows from equation (17). Now the c_{ij}^* in equation (18) are equal to $F_{ij+}^*/F_{+j+}^* = (\gamma\ \gamma_i^C\ \gamma_j^G)/(\gamma_j^G\ \Sigma_i\ \gamma_i^C) = \gamma\ \gamma_i^C/(\Sigma_i\gamma_i^C)$, which does not depend on j, so $c_{ij}^* = c_i^*$ is the standard set of weights and the $r_{j(k)}^*$ obtained from marginal CG purging are thus equivalent to direct-standardized rates. (Marginal CG purging with a particular standard group is equivalent to direct standardization also, as can be verified by arguments similar to those above.)

To summarize, the marginal-CG purging method is equivalent to the direct method of standardization, which is the most commonly used method of rate adjustment in sociology, demography, and other areas. The standard set of weights (c_i^*) implicit in the method of marginal-CG purging will usually differ from common choices for the weights in conventional standardization. For example, the c_i^* above will differ in general from the arithmetic average of the observed C marginal across groups. Of course, if a standard group is selected, the c_i^* are the same for both methods. Further relationships among methods are discussed in Clogg, Shockey, and Eliason (1987).

GROUP COMPARISONS OF ADJUSTED OR CRUDE RATES

Absolute and Relative Contrasts

If just two groups are compared, ordinarily one of two types of comparisons is of interest, the *arithmetic* (or absolute) contrast,

$A = r_{1(k)}^* - r_{2(k)}^*$, (simple differences in rates) or the *relative* comparison, $r_{1(k)}^*/r_{2(k)}^*$, sometimes called a relative-risk ratio (see Fleiss 1981). The value of A would be

zero under the case of equivalent rates. The relative-risk ratio would be one (or logarithm = 0) under the same condition. Relative comparisons thus can be made in terms of contrasts of log-rates,

$$L = \log(r\overset{*}{.}_{1(k)})$$
$$- \log(r\overset{*}{.}_{2(k)}).$$

The relative-risk ratio is $\exp(L)$. Still another way to make group comparisons is based on odds or logits. If the dependent variable is dichotomous, for example, we could compare $O_1 = r\overset{*}{.}_{1(1)}/r\overset{*}{.}_{1(2)}$ with $O_2 = r\overset{*}{.}_{2(1)}/r\overset{*}{.}_{2(2)}$, either with O_1/O_2 or $\log(O_1) - \log(O_2)$. (The procedures described below could be used to examine comparisons of odds or logits across groups, although we do not pursue this in any detail.)

For $J \geq 2$ groups, general linear contrasts of the form

$$A_k = \sum_{j=1}^{J} d_j \, r\overset{*}{.}_{j(k)} \qquad (19)$$

for absolute comparisons and

$$L_k = \sum_{j=1}^{J} d_j \log(r\overset{*}{.}_{j(k)}) \qquad (20)$$

for relative comparisons can be used, where $\Sigma_j \, d_j = 0$. For the two-group case above, $d_1 = +1$ and $d_2 = -1$. See Ott (1984) for discussion of linear contrasts and their utility for group comparisons in the context of analysis of variance. The idea is that quite specific group comparisons can be made in terms of equations (19) or (20). To see the usefulness of these quantities, situations involving more than two groups and more than two levels of the dependent variable should be considered. With just $J = 3$ groups, for example, linear contrasts with d values (1, -1, 0), (1, 0, -1), and (0, 1, -1) compare each pair of groups; linear trend can be examined in cases where the grouping variable represents a quantitative characteristic, and so forth. Contrasts of this general kind were used to summarize group differences (parity differences and age differences) in Tables 3 and 4. For multiple groups and multiple categories of the dependent variable, linear contrasts such as these allow for much more sensitive analysis of group differences

than is generally possible with the analyses of parameter values in log-linear models.

The Precision of Adjusted Rates and Contrasts Among Them

Except for some special cases, precision of adjusted rates has not been discussed in the literature on standardization or rate adjustment. In some instances, of course, the statistical tests that are practically automatic with log-linear analysis obviate the need for estimates of precision, particularly if the only goal is to test a null hypothesis of no group differences. For example, the model of conditional independence between G and D given C (hypothesis H_2 given earlier) is often a natural hypothesis to test. The chi-squared (goodness-of-fit) statistics associated with this model can be used to test the significance of group differences in the levels of the dependent variable (adjusted for compositional differences between the groups). Various kinds of restricted log-linear models (Haberman 1979; Goodman 1984) can also be used to obtain quite specific hypothesis tests of special interest in particular situations. However, statistical tests based on chi-squared statistics do not provide direct evidence about group differences in adjusted rates. A general method for calculating standard errors of adjusted rates and related quantities is required for both specific hypothesis tests (e.g., testing equality of $r\overset{*}{.}_{1(1)}$ and $r\overset{*}{.}_{2(1)}$) and interval estimation (e.g., an interval estimate for $r\overset{*}{.}_{1(1)} - r\overset{*}{.}_{2(1)}$ or for $\log(r\overset{*}{.}_{1(1)}/r\overset{*}{.}_{2(1)})$.

Let $s(.)$ denote the estimate of the standard error of a given quantity. We could use $s(r_{.j(k)},)$ $s(r\overset{*}{.}_{j(k)})$, $s(A_k)$, and $s(L_k)$ to calculate interval estimates of crude rates, adjusted rates, absolute contrasts of rates, and relative contrasts of rates, respectively, using large sample approximations. In many cases these interval estimates can be used to test null hypotheses of various kinds, merely by inspecting whether the null value is included in the interval. For absolute contrasts, the $(1 - \alpha)$ confidence interval is just

$$\hat{A}_k \pm \; z_{1-\alpha/2} \, s(\hat{A}_k), \qquad (21)$$

with similar expressions for any other quantity for which an interval estimate is desired, where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ percentile value of the standard unit normal distribution. (For

$\alpha = .05$, the z-value is 1.96.) The goal is to create estimates of standard errors for all possible quantities that might be of interest, for situations where both unsaturated or saturated models are used to describe the cell frequencies, and for all of the possible purging methods considered. For some cases, standard procedures (such as the delta method discussed in Bishop, Fienberg, and Holland 1975) can be used to derive the standard errors of interest. In other cases (e.g., where there are two groups and only two categories of the dependent variable), standard errors can be obtained from the standard errors for model parameters. But it is difficult to use such classical tools to derive the standard errors for all possible situations. The jack-knife technique, which is very easy to apply to contingency-table problems, is a general solution to the problem.

The jackknife technique is one of several "sample reuse" methods that can be used to calculate standard errors of parameter estimates. See Miller (1974), Efron (1979), and Mosteller and Tukey (1977) for details. Henry (1981) appears to be the first to apply jackknife methods in contingency-table problems; see also Fay (1985). Let θ denote a parameter of interest and $\hat{\theta}$ a statistic that estimates θ in a sample of size n. Let $\hat{\theta}_{(h)}$ denote the value of the statistic when the hth observation is deleted. The sample is "reused" n times to create estimates $\hat{\theta}_{(h)}$, $h = 1, \ldots, n$. In general the jackknife method requires n separate calculations. Define $\hat{\theta}_{(.)}$ as the mean of the $\hat{\theta}_{(h)}$,

$$\hat{\theta}_{(.)} = \left[\sum_{h=1}^{n} \hat{\theta}_{(h)} \right] / n. \qquad (22)$$

The estimated sampling variance of $\hat{\theta}$ or (or of $\hat{\theta}_{(.)}$) is

$$s^2(\hat{\theta}) = [n-1)/n] \sum_{h=1}^{n} \left[\hat{\theta}_{(h)} - \hat{\theta}_{(.)} \right]^2, \quad (23)$$

and $s(\hat{\theta}) = [s^2(\hat{\theta})]^{1/2}$ is the estimated standard error. The factor $(n-1)/n$ is of no consequence for large n; note that the variance estimator is essentially the sum of squares about the mean.

Now consider the three-way contingency table with cell frequencies f_{ijk}, and let $M = IJK$ denote the dimension of the contingency table (number of cells). Let θ denote a quantity such as a crude rate, an adjusted rate (obtained from any of the purging methods discussed here or from other methods), a log-rate, or the value of an absolute or relative contrast of rates. Let $\hat{\theta}$ denote the corresponding maximum likelihood estimate of θ based on the $\hat{F}_{ijk}$, the estimated expected frequencies under some model (saturated or unsaturated). Next define $\hat{\theta}_{(ijk)}$ as the estimate of θ when *one observation* in cell (i,j,k) is deleted. In other words, f_{ijk} is replaced by $f_{ijk} - 1$ and the estimate of θ is recalculated. (Note that we require $f_{ijk} \geq 1$, but the procedure can be modified when this condition is not satisfied.[2]) An important fact is that only one recalculation of the parameter estimate is required per cell. The value of $\hat{\theta}_{(ijk)}$ is the same as $\hat{\theta}_{(h)}$ for all f_{ijk} individuals in the (i,j,k) cell. The jackknife estimator of θ is now

$$\hat{\theta}_{(.)} = \left[\sum_{i,j,k} f_{ijk} \, \hat{\theta}_{(ijk)} \right] / n, \qquad (24)$$

and the estimated sampling variance of $\hat{\theta}$, or of $\hat{\theta}_{(.)}$, is

$$s^2(\hat{\theta}) = [(n-1)/n] \sum_{i,j,k} f_{ijk} \, [\hat{\theta}_{(ijk)} - \hat{\theta}_{(.)}]^2. \qquad (25)$$

The square root of $s^2(\hat{\theta})$ is the estimated standard error of $\hat{\theta}$. Note that M rather than n

---

[2] The jackknife method requires modification if one or more of the observed frequencies are zero. If only a few of the observed cell counts are zero, the easiest solution is to simply bypass the zero cells in the jackknife calculations; the calculations would then be carried out on $M - M^*$ cells, where M^* is the number of cells with zero entries. Of course, this situation should lead one to question the suitability of any *large-sample* approximation, including jackknife results. If there were many zero frequencies, for example, it would not be the case that the jackknife estimators calculated in the above manner would have the indicated normal distributions used for both hypothesis tests and interval estimates. In such cases, the user might be better off relying on chi-squared statistics for appropriately chosen log-linear models (such as H_1, H_2, or nested comparisons of them, as defined in the text). Note that zero frequencies create existence problems whenever the saturated model is used as a basis for purging. The maximum-likelihood estimates of the corresponding tau parameters do not exist (take on the boundary values zero or infinity) when one or more sampling zeroes occur. In this case the simplest solution is to add a constant such as one-half to the cell frequencies prior to estimating the parameters. See Clogg, Shockey, and Eliason (1987) for further discussion.

separate calculations are required, or the computational intensity of the procedure varies with the size of the contingency table, not with sample size. This procedure has been implemented in the computer program used to generate the results in Tables 3 and 4. It is important to realize that equation (25) is appropriate for any of the relevant quantities obtained from a rate adjustment procedure, such as the adjusted rates themselves or contrasts among them. (This procedure could also be implemented with the somewhat different methods given in Kaufman and Schervish (1986) or with the substantially different methods given in Davis (1984).) It should be noted that the standard errors calculated from the jackknife technique, and inferences based on them, are justified by the same general large-sample theory used to justify inferential procedures in popular statistical models for categorical data. For further details, see Clogg, Shockey, and Eliason (1987).

FURTHER ANALYSIS OF THE DATA IN TABLE 1

Four different purging methods were applied to the data in Table 1: the partial-*CG*, partial-*CG*-and-*CGD*, marginal-*CG*, and marginal-*CG*-and-*CGD* methods discussed above. Because there is strong evidence for three-factor interaction (equivalently, two-

factor age × parity interaction in the logit model), the saturated model was used for all calculations in this case. Results appear in Table 5; quantities in column (2) pertaining to the partial-*CG*-and-*CGD* method were used in Table 3.

The first two methods lead to substantially similar results, so the three-factor interaction does not have much of an effect on summary rates. This justifies summarizing the results using the partial-*CG*-and-*CGD* method, as was done in Table 3. There is of course some loss of information, but the two sets of quantities allow clear-cut description of the effect of three-factor interaction on the inferences. For example, the partial-*CG* method gives an adjusted difference between parity groups 1 and 2 of 39.4 percent, compared to 37.6 percent when the three-factor interaction is also purged, so the effect of three-factor interaction is about 1.8 percent (absolute) on this contrast.

The second two methods (see columns [3] and [4]) are also in close agreement, so once again three-factor interaction does not seem so perverse that summary rates are misleading. (An exception to this generalization is that the contrast between parity groups 2 and 3 is significant with marginal-*CG* purging but not with marginal-*CG*-and-*CGD* purging. However, most of the crude difference [9.6 percent] is removed with either method.) The salient feature of the results in Table 5 is the

Table 5. Adjusted Percentage Desiring More Children by Parity Group: Results From Four Purging Methods

| Parity Group | Purging Method | | | |
	Partial *CG*[a] (1)	Partial *CG* and *CGD*[b] (2)	Marginal *CG*[c] (3)	Marginal *CG* and *CGD*[d] (4)
1	60.0%	57.7%	53.6%	52.9%
2	20.6	20.1	22.7	21.8
3	17.5	18.2	18.1	18.9
4+	15.6	16.9	16.3	17.6
Contrasts:[e]				
1 vs. 2	39.4	37.6	30.9	31.1
	(3.3)	(4.1)	(2.2)	(3.0)
2 vs. 3	3.0	1.9	4.6	3.0
	(2.3)	(2.9)	(1.8)	(2.3)
3 vs. 4+	2.0	1.2	1.8	1.3
	(2.2)	(2.6)	(2.3)	(2.4)
1 vs. 4+	44.4	40.8	37.3	35.3
	(3.2)	(3.8)	(2.7)	(3.1)

[a] Purged of partial *CG* interaction; see (12).
[b] Purged of partial *CG* interaction and three-factor interaction; see (13).
[c] Purged of marginal *CG* interaction (equivalent to direct standardization); see (14).
[d] Purged of marginal *CG* interaction and three-factor interaction; see (15).
[e] Simple differences in percentages, parity 1 minus parity 2, etc. Standard errors of contrasts in parentheses.

difference between marginal-CG purging and partial-CG purging, not the difference between two-factor (CG) purging and purging that also takes account of three-factor interaction. But even here, the main difference is in the contrast between parity groups 1 and 2, with the partial-CG method giving an estimated difference of about 39 percent, while the marginal CG method gives an estimated difference of about 31 percent. In this case, the researcher must choose between the two basic approaches to adjusting for composition-group interaction (marginal or partial adjustment), recognizing the different rationale for each procedure. In general, it would be prudent to compare results from different methods to examine the dependence of the inferences on the method used. Note that there is no basic change in the inference for contrasts between parity groups 2 and 3 or between parity groups 3 and 4 for either method. The only problem is determining whether the adjusted difference between parity groups 1 and 2 should be estimated at around 39 percent or around 31 percent (where the unadjusted difference is 46 percent). In this case, it seems clear that no fundamental change in the substantive inference would result from using either figure. Similar comments apply when age differences in the desire for more children are adjusted for parity composition (as in Table 4), except there are even fewer differences among methods implying that the adjusted age differences are not dependent on the method of purging used.

In our opinion, it is useful to consider results from alternative purging methods to come to grips with the uncertainties associated with the adjustment process. Differences among purging methods are to be expected in some cases for many of the same reasons that differences among standardized rates depend (in some cases) on the standard set of weights used to make comparisons. However, we believe that the description of the data provided in Tables 3 and 5 is much more clear-cut than a description based on model parameter values in Table 2, even though there is some uncertainty in assigning a single value to the adjusted difference between parity groups 1 and 2.

DISCUSSION

The procedures presented here are quite general and can be used in many situations where a simplified interpretation of multivariate relationships among discrete variables is sought. It should be emphasized that the purging method and rate-adjustment procedures based on it do not replace formal statistical models for categorical variables. In fact, the purging method is based explicitly on the log-linear model for the contingency tables in question. The method of rate adjustment can be viewed as a translation of the parameter values of this model into quantities that might be more readily interpretable. While we have stressed the use of adjusted rates and absolute contrasts among them (such as simple group differences in the rates) — appealing to the inherent simplicity of rates or percentages — the procedures can be used to summarize group differences in log-rates or to analyze relative risk ratios and related quantities. A computer program available from the authors allows consideration of all of the urging methods described here, including the method of direct standardization, and uses the jackknife method for calculating the precision of relevant quantities. Group contrasts of rates, log-rates, and other quantities can be analyzed with this program.

Although the example is a three-contingency table with only three explicit variables, the methods are actually fully multivariate. Suppose that Y is the dependent variable and $X_1, \ldots, X_q$ are specified explanatory variables ($q \geq 2$), all of which are discrete. To study the effect of X_1 on Y adjusted for $X_2 - X_q$, let the levels of X_1 serve as the groups (i.e., $X_1 = G$). The methods of this paper can be applied in a straightforward manner with $X_2 - X_q$ combined to form a single (joint) categorical variable C. In other words, the composition variable C is the cross-classification of X_2, $X_3, \ldots, X_q$ when the effects of X_1 are singled out for special attention. To study the effect of X_2 on Y adjusted for X_1 and $X_3 - X_q$, define X_2 as the group variable and let the composition variable C be defined as the single (joint) variable obtained by cross-classifying $X_1, X_3, \ldots, X_q$. This procedure is repeated for each particular explanatory variable, regarding it as the group variable G and the rest of the explanatory variables (taken as a single joint variable) as the composition variable. The purging process would be implemented with the expected frequencies obtained from the particular

model chosen for the multivariate relationship. Examples with more than three variables appear in Clogg, Shockey, and Eliason (1987). In short, the rate-adjustment procedure considered here can be used to interpret multivariate relationships in practically all cases where logit models or logistic-regression models (for ordered or nominal-level dependent variables) are used. In cases where continuous predictors are used, the researcher might consider grouping them in some fashion to apply the method. The reader is invited to compare the present method with other methods for summarizing relationships estimated with logistic regression (Hanusheck and Jackson 1977).

REFERENCES

Allison, Paul D. 1984. *Event History Analysis.* Beverly Hills, CA: Sage Publications.

Altham, Patricia M.E. 1984. "Improving the Precision of Estimation by Fitting a Model." *Journal of the Royal Statistical Society, Series B* 46:118–9.

Anderson, Sharon, Arianne Augier, Walter W. Hauck, David Oakes, Walter Vandaele, and Herbert I. Weisberg. 1980. *Statistical Methods for Comparative Studies: Techniques for Bias Reduction.* New York: Wiley.

Bishop, Yvonne M.M. 1969. "Full Contingency Tables, Logits, and Split Contingency Tables." *Biometrics* 25:383–400.

Bishop, Yvonne M.M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: MIT Press.

Burch, Thomas K. and Ashok K. Madan. 1986. "A Note on the Components of Coale's I_g and Other Indirectly Standardized Indices." *Canadian Studies in Population* 13:151–66.

Clogg, Clifford C. 1978. "Adjustment of Rates Using Multiplicative Models." *Demography* 15:523–39.

———. 1979. *Measuring Underemployment: Demographic Indicators for the United States.* New York: Academic Press.

Clogg, Clifford C. and Scott R. Eliason. 1987. "Some Common Problems in Log-Linear Analysis." *Sociological Methods and Research* 15:4–44.

Clogg, Clifford C. and James W. Shockey. 1985. "The Effect of Changing Demographic Composition on Recent Trends in Underemployment." *Demography* 22:395–414.

———. 1988. "Multivariate Analysis of Discrete Data." In *Handbook of Multivariate Experimental Psychology,* edited by John R. Nesselroade and Raymond B. Cattell. New York: Plenum Press.

Clogg, Clifford C., James W. Shockey, and Scott R. Eliason. 1987. "A General Statistical Framework for Adjustment of Rates." Unpublished manuscript.

Das Gupta, Prithwis. 1978. "A General Method of Decomposing a Difference Between Two Rates into Several Components." *Demography* 15:99–112.

Davis, James A. 1984. "Extending Rosenberg's Technique for Standardizing Percentage Tables." *Social Forces* 62:679–708.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics* 7:1–26.

Fay, Robert E. 1985. "A Jackknifed Chi-Squared Test for Complex Samples." *Journal of the American Statistical Association* 80:148–57.

Fienberg, Stephen E. 1980. *The Analysis of Cross-Classified Categorical Data,* 2nd ed. Cambridge, MA: MIT Press.

Fleiss, Joseph L. 1981. *Statistical Methods for Rates and Proportions,* 2nd ed. New York: Wiley Press.

Fox, John. 1987. "Effect Displays for Generalized Linear Models." In *Sociological Methodology 1987,* edited by C.C. Clogg. Washington, DC: American Sociological Association.

Freeman, Daniel H., Jr. and Theodore R. Holford. 1980. "Summary Rates." *Biometrics* 36:195–205.

Goodman, Leo A. 1978. *Analyzing Qualitative/Categorical Data.* Cambridge, MA: Abt Books.

———. 1984. *The Analysis of Cross-Classifications Having Ordered Categories.* Cambridge, MA: Harvard University Press.

Grizzle, James E., Charles F. Starmer, and Gary G. Koch. 1969. "Analysis of Categorical Data by Linear Models." *Biometrics* 25:489–504.

Haberman, Shelby J. 1978. *Analysis of Qualitative Data, vol. 1: Introductory Topics.* New York: Academic Press.

———. 1979. *Analysis of Qualitative Data. vol. 2. New Developments.* New York: Academic Press.

Hanushek, Eric A. and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.

Henry, Neil. 1981. "Jackknifing Measures of Association." *Sociological Methods and Research* 10:233–40.

Kaufman, Robert L. and Paul G. Schervish. 1986. "Using Adjusted Crosstabulations to Interpret Log-Linear Relationships." *American Sociological Review* 51:717–33.

Keppel, Kenneth G. 1981. "Mortality Differentials by Size of Place and Sex in Pennsylvania for 1960 and 1970." *Social Biology* 28:41–8.

Keyfitz, Nathan. 1966a. "Sampling Variance of Demographic Characteristics." *Human Biology* 38:309–16.

———. 1966b. "Sampling Variance of Standardized Mortality Rates." *Theoretical Population Biology* 38:309–16.

———. 1968. *Introduction to the Mathematics of Population.* Reading, MA: Addison-Wesley.

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50:1168–94.

———. 1964. "Standardized Comparisons in Demographic Research." *Demography* 1:296–315.

———. 1966. "Theoretical Considerations in the Selection of a Mortality Index, and Some Empirical Comparisons." *Theoretical Population Biology* 38:293–308.

Little, Roderick J.A. and Thomas W. Pullum. 1979. "The General Linear Model and Standardization: A Comparison." *Sociological Methods and Research* 7:475–501.

Miller, Rupert G. 1974. "The Jackknife—A Review." *Biometrika* 61:1–15.

Mosteller, Frederick and John W. Tukey. 1977. *Data Analysis and Regression.* Reading, MA: Addison-Wesley.

Ott, Lyman. 1984. *An Introduction to Statistical Methods and Data Analysis*, 2nd ed. Boston: Duxbury Press.

Rindfuss, Ronald R. and Larry L. Bumpass. 1978. "Age and the Sociology of Fertility: How Old is Too Old?" Pp. 43–56 in *Social Demography*, edited by K.E. Taueber, L.L. Bumpass, and J.A. Sweet. New York: Academic Press.

Rosenberg, Morris. 1962. "Test Factor Standardization as a Method of Interpretation." *Social Forces* 41:53–61.

Shah, Mohammed R. 1986. *Standardization in Demographic Analysis.* Ph.D. Thesis, Inter-University Program in Demography, Vrije Universiteit Brussels, Belgium.

Shryock, Henry S. and Jacob S. Siegel. 1980. *The Methods and Materials of Demography.* (two volumes, fourth printing, rev.) Washington, DC: U.S. Government Printing Office.

Smith, Kent W. 1977. "Test-Factor Standardization and Marginal Standardization." *Social Forces* 56:240–9.

Spiegelman, Mortimer. 1968. *Introduction to Demography*, rev. ed. Cambridge, MA: Harvard University Press.

Stinchcombe, Arthur L. 1968. *Constructing Social Theories.* New York: Harcourt, Brace, and World.

Whittemore, Alice S. 1978. "Collapsibility of Multidimensional Contingency Tables." *Journal of the Royal Statistical Society, Series B* 40:328–40.

Wunsch, Guillame J. and Marc G. Termote. 1978. *Introduction to Demographic Analysis: Principles and Methods.* New York: Plenum Press.

RESEARCH NOTES

◆

CLASS IDENTIFICATION PROCESSES OF MARRIED, WORKING MEN AND WOMEN*

IDA HARPER SIMPSON   DAVID STARK
Duke University   University of Wisconsin-Madison

ROBERT A. JACKSON
Duke University

We propose gender-specific models of class identification based on the distinctive work and occupational experiences of men and women. Using data from the General Social Surveys of the National Opinion Research Center on white married men and women who were employed full time in 1973, 1974, 1976, 1977, and 1980, regression analysis supports our contention that processes of class identification are gender-specific. Men's class identification is predicted by whether they do manual or nonmanual work and by their position in a supervisory hierarchy. Women's is predicted by self-employment, work in a female occupation, and union membership.

Men and women differ in their class identification processes (Vanneman and Pampel 1977; Robinson and Kelley 1979; Simpson and Mutran 1981). Factors that explain men's class identification do not predict women's as well, yet there has been little systematic study of this gender difference. Instead, emphasis has been on explaining either men's or women's class identification or one's spouse's contribution to the class identification of the other (e.g., Jackman and Jackman 1983). Similarities and differences between the sexes have not been studied as such. This paper deals with this neglected topic.

For men, there are two traditions of class-identification studies: the structural tradition and the socioeconomic-status tradition. The structural tradition sees class identification as shaped by relation of control and authority within the workplace (e.g., Robinson and Kelley 1979; Vanneman and Pampel 1977; Wright 1978). The socioeconomic tradition sees class identification as resting on societal statuses, especially occupation, edu-

cation, and income (e.g., Hodge and Treiman 1968; Jackman and Jackman 1983). Both traditions locate the main influences on men's class identification in their work worlds or in social achievements related to work.

For women, there are three traditions of studies. One is the same structural approach as has been applied to men (Robinson and Kelley 1979). The other two are variants of the socioeconomic-status tradition and focus on married women. One, a familial explanation, asserts that wives' class identification is based on their husbands' statuses; the other argues that wives' own autonomous socioeconomic statuses also affect their class identification (e.g., Ritter and Hargens 1975). The latter explanation, which was advanced by considerable research in the 1970s, was reversed in favor of the former by Jackman and Jackman (1983). We re-examine Jackman and Jackman's findings and consider them along with earlier findings they dispute.

From a review of these socioeconomic studies, we draw two main inferences to inform our thinking on married women's class identification. A familial model of class identification applies to nonworking wives; they derive their class identification from their husbands' statuses of occupational prestige, education, and family income (Simpson and Mutran 1981; Jackman and Jackman 1983). But working wives do not draw as heavily on their husbands' statuses;

working women take into account their own autonomous statuses (Ritter and Hargens 1975). The familial explanation may well have applied before it was normative for wives to have continuous paid employment, but, increasingly, women's labor-force participation resembles men's (Hartmann 1987), and their societal roles are increasingly independent of their husbands' socioeconomic statuses. When wives participate autonomously in the society in ways that influence how others treat them in class terms, as in the workplace, their autonomous participation affects their class identification. Their husbands' statuses may also affect their class identification because wives continue to share family resources and because husbands continue to be assigned the quasi-legal status of head of the household. But husbands' influences are not sufficient to justify a familial model that ignores wives' autonomous work roles and experiences.

Our second inference from earlier studies is that models based on autonomous socioeconomic factors do not predict working wives' class identifications as well as husbands' (Simpson and Mutran 1981; Jackman and Jackman 1983). Just as familial models do not fully capture the influences that shape working wives' class identification, neither do socioeconomic models based on autonomous socioeconomic statuses. Employed wives differ both from nonemployed wives and from husbands in their class identification. They differ from nonemployed wives in that they have autonomous roles that affect their class identification and they differ from husbands in that the autonomous socioeconomic models do not capture as fully the influences that shape their class identification.

Our review of earlier research leads us to disagree with one of Jackman and Jackman's (1983) major conclusions in *Class Awareness in the United States*. They argue that the earlier research that questioned wives' status borrowing from husbands (e.g., Ritter and Hargens 1975) failed to control wives' education, which, when controlled, shows that wives draw on their husbands' rather than their own occupations. They conclude that "the family appears to be the unit of stratification. Within the family, the husband is the principal source of status, and this state of affairs is unaffected by the wife's labor force participation" (1983, p. 164). In other words, an autonomous model is appropriate

for working husbands, but a familial one composed of husbands' statuses is appropriate for working wives. Thus, Jackman and Jackman hold that wives' labor-force participation has no effect on their class identification. This too is contrary to what we are proposing. We argue that wives' work is important, and that the socioeconomic model, which Jackman and Jackman use, fails to take into account work experiences that bear on wives' class identification. We argue that wives' work experiences affect their class identification and that models should include their autonomous statuses and experiences just as they include them for husbands. But autonomous models made up of only socioeconomic factors (education, occupational prestige, and income) do not fit working wives as well as husbands.

Jackman and Jackman used models developed to study men as standards to observe women (cf. Acker 1980; Feldberg and Glenn 1979). Other studies that have examined the effects of workplace variables on class identification have not focused on gender, but they have nonetheless uncovered gender differences in class identification (e.g., Robinson and Kelley 1979; Vanneman and Pampel 1977). Workplace variables seem to affect both men's and women's class identification, but in different ways because their work experiences differ. Drawing from these structural studies, our study proposes gender-specific processes of class identification. We hypothesize that work processes shape employed wives' class identification as well as husbands', but that the way in which this happens differs.

GENDER, WORK, AND CLASS IDENTIFICATION

When men and women identify with a class, they see themselves as belonging to it. Class identification expresses in a symbolic way their own experiences with power and privileges (Lockwood 1966). We assume that some of the most profound experiences with power and privileges take place at work or in conjunction with it and that these experiences are conditioned by gender. We propose gender-specific models that view class identification as derived from workplace and occupational dimensions of work. The *male model* derives class identification from four variables: manual/nonmanual work classifica-

tion; authority relations; self-employment; and union membership. Doing manual work, being supervised, being an employee, and belonging to a union beget a working-class identification. Our *female model* predicts class identification from one distinctive variable — sex segregation — and from self-employment and union membership, which are also in the male model. Work in a female-typed occupation, sale of one's labor (as opposed to self-employment), and union membership promote women's identification with the working class. The manual/nonmanual skill classification and the authority dimension, which we predict to affect men's class identification, we do not think will influence women's, and sex-segregation, which we think will affect women's class identification, we do not think will influence men's.

In proposing these gender-specific models, we assume that the key dimensions of male jobs may not be present in the same form in female ones and vice versa, and that the connotations of working-class membership are not necessarily the same for men and women. The manual/nonmanual distinction that is important for men is not as salient a class dimension for women (Vanneman and Pampel 1977), in large part because the nonmanual and manual jobs women hold differ from those held by men. Both the manual and nonmanual jobs men hold are less open to women (Oppenheimer 1970; Form 1985). For example, though women have long worked in manual jobs in service and nondurable goods industries, their work requires stamina and dexterity more than physical strength. This work is less linked to the male working-class image, an image that connotes physical exertion. Jobs in nondurable goods industries, such as in canneries, are less regular than in durable goods industries (see U.S. Bureau of Census 1973, Table 11), and young women, in particular, leave them for lower white-collar and service jobs. Their seasonality and worker turnover impair identification with the working class. Conversely, we do not think occupational sex segregation affects men's class identity. Men are infrequently excluded from occupations, even female ones (Oppenheimer 1970). Their experiences with social exclusion, unlike women's, have not been based on gender; instead, men react more to subordination to authority. Authority relations for women are

less linked to class identification (Robinson and Kelley 1979) because women tend to work in smaller groups where authority is exercised more directly, more personally, and in paternalistic ways (e.g., the secretary and her boss; the nurse and the doctor; the teacher and her principal; the waitress and the cook). In such work groups, authority tends to foster compliance with supervision rather than organized opposition to it.

We predict that self-employment promotes identification with the middle class (Jackman and Jackman 1983), but in somewhat different ways for men and women. Self-employment provides autonomy and property rights to one's labor (Robinson and Kelley 1979), qualities that are identified with the middle class. Male self-employment cross-cuts manual and nonmanual work (Form 1982), but women's is concentrated in nonmanual and service occupations. In 1970, 48 percent of self-employed men held nonmanual and service jobs (excluding domestic) compared with 83 percent of self-employed women, and in 1980 the percentages were 52 and 84 respectively. Women's self-employment is typically in female-typed occupations, where it serves as an outlet from the blocked mobility characteristics of those occupations. Men want to be their own bosses (Chinoy 1955); a correlation of self-employment with class identification for men would be mediated in part through supervision. This counteracting effect of supervision on self-employment is probably less likely to occur among women, whose work settings are more personalized than men's. Women, however, may orient as much as men to the independence of self-employment and to its opportunities for higher earnings (Form 1985). We expect union membership to affect men and women in essentially the same way, because unions are highly identified with the working class, even though the kinds of jobs held by male union members tend to be closed to women and female union members are concentrated in female service work (Baron and Bielby 1982; Form 1985).

In testing for these relationships, we control socioeconomic statuses, on which we have data, that earlier research was found to influence class identification. These are education and occupational prestige of the respondent and spouse, and family income. (Jackman and Jackman included respondent's as well as spouse's income, but since our data

set does not include a spouse-income question, we use family income instead. With this exception we represent fully the variables in the Jackmans' socioeconomic model as controls for our work- and occupationally based models.) We expect our models to predict class identification with the socioeconomic statuses controlled.

SAMPLE, DATA, AND METHODS

We restrict our sample to employed, white, married men and women to concentrate on gender without confounding it with racial or marital influences. Jackman and Jackman (1983) found significant racial variation in the effects of socioeconomic factors on class identification, and other research has shown similar racial variation in marital power and in the occupational attributes of husbands and wives. Together these earlier findings suggest that the familial model is more likely to apply to whites than to blacks. In limiting our sample only to whites, we choose the harder test; in white marriages more than in black marriages, husbands' statuses exceed their wives, inclining wives to "borrow status from their husbands."

We impose the marital restriction because the family situations of unmarried women and men do not fit the assumption that we are examining; similarly the nonemployed lack autonomous work statuses and experiences. We exclude the unemployed because there are too few in our data set and also because our workplace variables do not apply to the unemployed.

With the exception of one measure to be discussed below, our data are taken from the 1973, 1974, 1976, 1977, and 1980 General Social Surveys of the National Opinion Research Center (Davis 1980). The surveys are on probability samples of the noninstitutional, voting-age population of the 48 contiguous states. We select for our study the full samples of married, white, full-time male and female participants in the labor force to test our gender-specific models. After determining that there was no significant yearly variation in the data for our measures, we merged the five surveys for each gender group to increase the sizes of the subsamples. The combined surveys yielded a total of 1,264 married, white, male workers and 577 married, white, female workers who met

selection criteria and who responded to the questions used in this research.

The dependent variable of our study is class identification. It is measured with the question, "If you were asked to use one of four names for your social class, which one would you say you belong to: the lower class, the working class, the middle class, or the upper class?" Since only 3.9 percent of male respondents and 2.2 percent of females chose the combined lower and upper classes, we deleted these respondents from our study. The distributions of class identification in the male and female subsamples are similar: 50.0 percent of the males and 54.0 percent of the females identified with the working class.

The independent variables in our male and female models are manual/nonmanual work classification, self-employment, supervisory relations, union membership, and occupational sex segregation. We use 1970 census designations to classify occupations as manual or nonmanual because most of the surveys were conducted in the 1970s. Four types of supervisory relations are designated by cross-classifying responses to the questions "Do you supervise anyone on your job?" and "Do you have a supervisor on the job to whom you are responsible?": (1) those who give orders only; (2) those who give and receive orders; (3) those who receive orders only; (4) those who are outside a supervisory system and neither receive nor give orders. Our measures of self-employment and union membership are both dichotomized: self-employed or not, belonged to a union or not. Sex segregation is defined on the basis of the percentage of females in the total labor force of each of the 441 occupations listed in the 1970 decennial census (U.S. Bureau of Census 1973, pp. 1–11). We classify occupations into three sex types using Burris and Wharton's scheme (1982): (1) occupations whose labor force had 85 percent or more females, called "female occupations"; (2) those in which females made up from 16 through 84 percent, termed "gender-mixed"; and (3) those in which males constituted 85 percent or more of the labor force are called "male occupations." (We use categorical definitions of sex-typing instead of a continuous measure, such as percent of an occupation female, because the theory of sex-typing presumes high concentration of a sex in an occupation.)

The variables in the socioeconomic model that we use as controls are education of

Table 1. Means and Standard Deviations for White, Full-Time Employed, Married Men and Women

Variables	Married Men (N=1,264)		Married Women (N=577)	
	$\bar{X}$	s.d	$\bar{X}$	s.d
Workplace and Occupational Variables				
Occupation: Manual = 1	.538	.499	.289	.454
Supervision:				
Gives orders = 1	.128	.334	.071	.257
Gives and receives orders = 1	.368	.482	.345	.476
Receives orders = 1	.434	.496	.516	.500
Outside hierarchy = 1	.070	.255	.068	.251
Self-employed: Self = 1	.146	.353	.069	.254
Sex segregation:				
Male occupation = 1	.688	.463	.090	.287
Gender mixed = 1	.305	.460	.567	.496
Female occupation = 1	.007	.084	.343	.475
Union membership: Member = 1	.313	.464	.153	.360
Socioeconomic Control Variables				
R's Hodge-Seigel-Rossi Prestige	42.398	13.901	42.286	13.375
Spouse's H-S-R Prestige	41.540	9.635	41.382	12.529
Respondent's education, in years	12.751	3.183	12.627	2.520
Spouse's education, in years	12.328	2.481	12.458	3.207
Spouse's labor force status:				
Out of labor force = 1	.525	.500	.107	.310
Family income, natural log	10.240	.549	10.237	.552
Dependent Variable				
Class identification:				
Working class = 1	.500	.500	.541	.499

respondent and of spouse, prestige of respondent's and spouse's occupation, spouse's labor-force status, and combined family income. Education is measured by number of years of schooling completed. Occupational prestige is measured by Hodge-Siegel-Rossi prestige scores. If the spouse was not currently employed (as many wives of working husbands were not), the code is "out of the labor force."[1] Family income is measured as logged dollar income. Dollar estimates are derived from the midpoints of the GSS income categories. Values for open-ended upper-income categories were estimated using Pareto curve techniques (Parker and Fenwick 1983). All years adjusted to 1981 dollars from consumer price index reports.

Table 1 gives the means and standard deviations of all variables used in the analysis. With the exception of the supervisory variables, variables in the occupational and workplace model are distributed differently among our male and female samples, while those in the socioeconomic model are similar. Men are disproportionately in male

occupations and women are mainly in gender-mixed ones. Women have about the same percentage of their labor force in female occupations as men have in gender-mixed ones (34.3 percent vs. 30.5 percent). Men are about twice as likely to be manual workers and to belong to unions as are women. Nonetheless, all variables are fairly well distributed in both the husbands' and wives' subsamples, with the exception of sex-typed occupations. Gender-mixed occupations show sizable proportions of both sexes, but the male and female occupational distributions are highly skewed. Nine percent of wives are in male occupations and only one-half percent of the males are in female occupations. The sex-typing of occupations limits the representation of the opposite sex in gender-typed occupations.

Table 2 gives the zero-order correlation coefficients for all variables for both wives and husbands. In our work and occupational model, shown in Quadrant A of Table 2, only five coefficients for husbands are greater than or equal to 0.30: self-employment and categories of supervision (.54, −.30, 0.33), manual occupation and union membership (.36), and manual occupation and male sex-typed occupation (.30). These correla-

Table 2. Zero-order Correlation Coefficients of Workplace, Occupational, and Socioeconomic Variables for Married, Working Men and Women (Females above the diagonal; males below)

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Occupation, respondent (Manual = 1)		-.08	-.16	.17	.04	.01	-.01	-.06	.07	.25	-.69	-.31	-.46	-.41	.06	-.35	.26
2. Supervision (Gives orders = 1)	-.15		-.29	-.29	-.07	.22	.10	-.07	.01	-.08	.09	.01	.03	.02	-.03	.03	-.07
3. Supervision (Gives & receives = 1)	-.13	-.20		-.75	-.28	-.16	-.11	.09	-.03	.18	.23	.15	.21	.17	.02	.15	-.13
4. Supervision (Receives orders = 1)	.24	-.34	-.67		-.20	-.24	-.06	.05	.04	.25	-.26	-.11	-.19	-.17	-.02	-.11	.17
5. Supervision (Outside hierarchy = 1)	-.02	-.10	-.21	-.24		.25	.32	-.04	-.02	-.10	-.02	-.07	-.05	.00	.04	-.09	-.03
6. Self-employed	-.12	.54	-.24	-.30	.33		.15	-.15	-.01	-.19	-.02	-.01	-.06	-.01	.04	-.06	-.10
7. Sex segregation (Male occupation = 1)	.30	.13	-.05	-.06	.04	.15		-.36	-.23	.00	.03	-.01	-.09	-.07	.07	-.04	.00
8. Sex segregation (Mixed occupation = 1)	-.29	.13	.05	.05	-.04	-.15	-.98		-.83	.09	.07	.08	.12	.07	-.06	.11	-.10
9. Sex segregation (Female occupation = 1)	-.05	-.03	-.01	.04	-.02	-.01	-.13	-.06		-.04	-.09	-.07	-.07	-.03	.02	-.09	.11
10. Union member	.36	-.17	-.08	.25	-.10	-.19	.02	-.01	-.04		-.19	-.14	-.10	-.12	-.01	.00	.17
11. R's Hodge-Seigel-Rossi Prestige	-.65	.12	.16	-.24	.01	.10	-.08	.08	.02	-.30		.40	.62	.43	-.02	.35	-.30
12. Spouse's H-S-R Prestige	-.25	.07	.07	-.10	-.03	.04	-.07	.06	.03	-.11	.40		.43	.52	-.01	.27	-.31
13. Respondent's education, in years	-.53	.08	.13	-.17	-.04	.00	-.18	.19	-.03	-.22	.58	.29		.58	-.09	.32	-.31
14. Spouse's education, in years	-.38	.07	.10	-.14	.00	.04	-.12	.12	.02	-.18	.40	.41	.60		-.11	.35	-.35
15. Spouse's LF status (Out of labor force = 1)	.08	.04	-.04	.01	.01	.05	.10	-.10	-.03	.04	-.01	-.02	-.05	-.15		-.27	.08
16. Family income, natural log	-.28	.17	.16	-.24	-.07	.11	-.01	.01	-.01	-.02	.29	.21	.31	.32	-.17		-.31
17. Subjective class ID (Working class = 1)	.42	-.16	-.13	.23	.01	-.11	.07	-.07	-.01	.20	-.39	-.16	-.39	-.33	.02	-.35	

(Quadrant A: rows 1–10, columns 1–10; Quadrant B: rows 1–10, columns 11–17; Quadrant C: rows 11–17, columns 1–10; Quadrant D: rows 11–17, columns 11–17.)

Notes: Quadrant A: Work- and occupationally based model correlations for men and women.
Quadrant B: Between-model correlations for women.
Quadrant C: Between-model correlations for men.
Quadrant D: Socioeconomic model correlations for men and women.

tions are as we expected. For wives only the correlation of self-employment and male-typed occupation exceeds 0.30. Manual work is associated with male-typed occupations and union membership for husbands, but for wives the variables appear fairly independent of each other.

In the socioeconomic status model, more variables are correlated with each other than in our work-based model, especially for wives (see Quadrant D). Five of the fifteen coefficients exceed .30 for men and eight do for women. Moderate to high correlations exist between respondent's and spouse's prestige and respondent's and spouse's education and between these four variables and family income for both husbands and wives. The correlations of prestige with education and with income are not surprising since they are components of some prestige measures; they caution us to be attentive to problems of multicollinearity in this model, particularly for women. Correlations between variables from the work-based and the socioeconomic models show the two sets of variables fairly independent of each other with one exception. (See Quadrant B for women; Quadrant C for men.) Manual work correlates substantially with respondent's prestige and education, and to a lesser extent with spouse's education, spouse's prestige, and family income. Other variables in our work-based model correlate weakly, if at all, with those in the socioeconomic model.[2]

On the whole, the patterns of association both within and between models are similar for wives and husbands, with the following notable exceptions. For our work-based model, manual work correlates with male occupations and union membership for husbands but not for wives. This is consistent with our view that men and women work in different occupational milieus that interrelate the variables in our model differently. In the socioeconomic-status model, the gender differences suggest that women garner socioeconomic capital from their spouses more than men; otherwise the coefficients are more similar than in our work-based model.

Our research is based on an assumption that husbands' and wives' work-based models differ. For a rigorous examination of this assumption, we conducted three statistical tests. We pooled the husbands' and wives' subsamples and tested the null hypothesis that the estimated parameters of the husbands' and wives' models are the same (Teberg 1986). This test for global differences of male and female parameter estimates yielded an F of 2.2915 (15 and 1811 d.f.) with a probability of 0.0033. Another more specific null hypothesis we tested was that the effects of workplace variables were the same for men and women. For this test, the effects of workplace predictors but not socioeconomic controls were allowed to vary by sex. This test showed an F of 2.7147 (9 and 1817 d.f.) with a probability of 0.0039. And last, in a combined model, we computed gender interaction terms for all variables. The interaction terms for manual occupations and self-employed were significant at better than the 0.01 and 0.05 levels respectively, and the union membership term was significant at the 0.06 level. But the interaction terms for the authority and sex-segregation measures were not significant.

We conclude from these results that there are firm statistical grounds for separating the total sample into gender subsamples and examining gender-specific models. The terms clearly suggest that manual occupations, self-employment, and union membership will predict differently for men and women within our gender-specific models. By comparison, the interaction terms for authority and sex segregation are not significant. But this does not imply that their effects *within* the gender models will be nonsignificant.

FINDINGS

Our first hypothesis is that autonomous models of class identification apply to working wives as well as husbands. Table 3 gives the regression of class identification on the variables in the socioeconomic and the work- and occupationally based models. Among the socioeconomic controls, we compare the influence on class identification of variables based on autonomous statuses with the influence of spousal statuses and family income. These controls replicate Jackman and Jackman's procedures.

For husbands, Table 3 shows that hus-

---

[2] Collinearity diagnostics do not indicate for specific variables any variance-inflation factor of critical magnitude. The overall condition index values for the male and female models are 3.44 and 3.74 respectively, well below a critical value. To the degree that collinearity is a problem, it is more apparent within the socioeconomic control model.

Table 3. Work-Related and Control Variable Predictors of Working-Class Identification Among White, Full-time Employed, Married Men and Women

Variables	Married Men (N = 1,264)			Married Women (N = 577)		
	Beta	b	s.e.	Beta	b	s.e.
Intercept		2.7806	.2503		2.8662	.3986
Workplace and Occupational Variables						
Occupation: Manual = 1	.1966	.1971***	.0361	−.0042	−.0047	.0586
Supervision:[a]						
Gives orders = 1	−.0809	−.1210*	.0491	−.0403	−.0782	.0780
Gives and receives orders = 1	−.0635	−.0659*	.0283	−.0513	−.0538	.0429
Outside hierarchy = 1	−.0205	−.0403	.0555	−.0399	−.0793	.0801
Self-employment: Self = 1	−.0090	−.0127	.0462	−.1063	−.2085*	.0820
Sex segregation:[b]						
Male occupation = 1	−.0143	−.0154	.0287	.0359	.0625	.0710
Female occupation = 1	−.0078	−.0465	.1450	.0827	.0868	.0409
Union membership: Member = 1	.0411	.0443	.0289	.1105	.1532**	.0549
Socioeconomic Control Variables						
R's Hodge-Seigel-Rossi Prestige	−.0711	−.0026*	.0013	−.0532	−.0020	.0022
Spouse's H-S-R Prestige	.0326	.0017	.0014	−.1090	−.0043*	.0018
Respondent's education, in years	−.1243	−.0195***	.0055	−.0630	−.0125	.0107
Spouse's education, in years	−.0908	−.0183***	.0065	−.1436	−.0223**	.0079
Spouse's labor force status:						
Out of labor force = 1	−.0438	−.0439	.0250	.0154	.0248	.0629
Family income, natural log	−.1940	−.1767***	.0249	−.1774	−.1605***	.0392
R^2		.2705			.2027	

[a] Omitted category is "receives orders."
[b] Omitted category is "mixed occupation."
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

bands' prestige and education are significant and more salient for them than are their wives' prestige and education, although wives' education is a significant predictor. For wives, the reverse holds, with spousal variables being more influential. Increments to husband's prestige and education significantly increase wives' middle-class identification, while increments to their own occupational prestige and education do not. These findings are generally consistent with those of Jackman and Jackman and their status-borrowing hypothesis, with two exceptions. For husbands, we find that increments to wives' education increase husbands' identification with the middle class (Simpson and Mutran 1981) and, unlike Jackman and Jackman's findings for wives, our findings show no significant independent association of wife's education with class identification with all the other variables controlled. Standard collinearity diagnostics indicate that education's lack of significance is due in part to its high multiple correlation with other socioeconomic variables. This is the only variable for which multicollinearity appears to affect interpretation of these findings.

Thus, net of Jackman and Jackman's socioeconomic model, which is given some credence by our findings, we wish to test our hypotheses. Findings in Table 3 also show that autonomous models of class identification can be applied to both working wives and husbands if workplace and occupational variables are included. We find that the autonomous statuses of union membership, self-employment, and work in a female occupation in our work-based model predict wives' class identification. These findings challenge Jackman and Jackman's conclusion that wives' work does not influence their class identification.

Our second hypothesis is that different work experiences are associated with husbands' and wives' class identification. The findings on our work- and occupationally based models given in Table 3 conform to this thesis. Husbands' identification with the working class is influenced by manual work and being supervised, while wives' is influenced by being a union member and an employee (as opposed to self-employment) and by work in a female occupation. Contrary to our expectations, self-employment and

union membership did not predict men's class identification. This finding on self-employment is contrary to Jackman and Jackman (1983) and our finding that union membership had no association net of other variables agrees with Hodge and Treiman (1968) and Jackman and Jackman (1973). There is no overlap of the variables significant beyond the .05 level for the two sexes.

Not only do the male and female models differ in the variables that influence class identification, but they also differ in the pattern of association of the variables with class identification. The central element in the male model is manual or nonmanual work, followed by authority position in the supervisory chain. (The betas are .197 for manual work and $-.081$ for giving orders.) But no one variable defines class identification as clearly for wives as manual work does for men. Instead, three occupation-related variables predict women's class identification in about the same degree: union membership (beta of .111), self-employment (beta of -0.106) and work in a female-typed occupation (beta of .083). The male and female models differ not only in the salient variables, but also in the degrees to which their constituent independent variables are related to class identification.

CONCLUSIONS

We draw three main conclusions from our research on the effects of workplace and occupational variables, net of socioeconomic statuses: (1) Working wives do not simply borrow their class identification from their husbands. Work experiences are a core part of the class-identification process of employed wives, as is true for husbands. Thus, a family model is inadequate to capture influences that shape class identification for either wives or husbands. Occupational and workplace variables predict both sexes' class identification independently of their own and their spouses' socioeconomic statuses. For employed wives, as well as for husbands, work statuses carry class meanings as much as or more than do familial statuses. This conclusion differs from Jackman and Jackman's. (2) The workplace and occupational variables related to class identification differ for men and women, without any overlap. Men's class identification is predicted by whether they do manual or nonmanual work and whether they are

subordinated in a chain of command. Women's is predicted by membership in unions, self-employment, and work in a female occupation. Working-class or middle-class identification means something very different to working women than to men and is likely to continue to have different meanings as long as men and women have different work experiences. (3) The class work of men is much more crystallized around the workplace than women's. Working-class identification for men is shaped substantially by manual work and by being in a chain of command in which orders have to be carried out. For women, working-class identification is not so concretized in the workplace. Rather, it is more occupationally based; women's work settings are more heterogeneous with categories of union membership or nonmembership, self-employment or employed work, and female or other occupations. Gender differences in the crystallization of class identification may reflect differences in the history of men's and women's work. The male pattern of class identification appears to have been institutionalized through industrialization, which defined work along manual and nonmanual lines and which instituted bureaucratized supervisory structures. Women's identification processes are developing out of their participation in a service economy that emphasizes location in occupationally related membership categories.

REFERENCES

Acker, Joan R. 1980. "Women and Stratification: A Review of Recent Literature." *Contemporary Sociology* 9:25–34.

Baron, James N. and William T. Bielby. 1982. "Workers and Machines: Dimensions and Determinants of Technical Relations in the Workplace." *American Sociological Review* 47:175–88.

Burris, Val and Amy Wharton. 1982. "Sex Segregation in the U.S. Labor Force." *Review of Radical Political Economics* 14:43–56.

Chinoy, Eli. 1955. *Automobile Workers and the American Dream*. Garden City, NY: Doubleday.

Davis, James A. 1980. *General Social Surveys, 1972–1980*. [MRDF]. Chicago: National Opinion Research Center [producer]. Storrs, CT: The Roper Center [distributor].

Feldberg, Roslyn L. and Evelyn Glenn. 1979. "Male and Female: Job Versus Gender Models in the Sociology of Work." *Social Problems* 26:524–38.

Form, William. 1982. "Self-Employed Manual Workers: Petty Bourgeois or Working Class?" *Social Forces* 60:1050–69.

———. 1985. *Divided We Stand*. Urbana: University of Illinois Press.

Hartmann, Heidi I. 1987. "Changes in Women's Economic and Family Roles in Post World War II United States." In *Women, Households, and the Economy*, edited by Lourdes Beneria and Catherine Stimpson. New Brunswick, NJ: Rutgers University Press.

Hodge, Robert W. and Donald J. Treiman. 1968. "Class Identification in the United States." *American Journal of Sociology* 73:535–47.

Jackman, Mary R. and Robert W. Jackman. 1973. "An Interpretation of the Relation Between Objective and Subjective Social Status." *American Sociological Review* 38:569–82.

———. 1983. *Class Awareness in the United States.* Berkeley: University of California Press.

Lockwood, David. 1966. "Sources of Variation in Working-Class Images of Society." *Sociological Review* 14:249–67.

Oppenheimer, Valerie Kincade. 1970. *The Female Labor Force in the United States.* Berkeley: Institute of International Studies, University of California.

Parker, Robert N. and Rudy Fenwick. 1983. "The Pareto Curve and Its Utility for Open-Ended Income Distributions in Survey Research." *Social Forces* 61:872–85.

Ritter, Kathleen V. and Lowell L. Hargens. 1975. "Occupational Positions and Class Identification: A Test of the Asymmetry Hypothesis." *American Journal of Sociology* 80:934–48.

Robinson, Robert V. and Jonathan Kelley. 1979. "Class as Conceived by Marx and Dahrendorf: Effects on Income Inequality and Politics in the United States and Great Britain." *American Sociological Review* 44:38–54.

Simpson, Ida Harper and Elizabeth Mutran. 1981. "Women's Social Consciousness: Sex or Worker Identity." Pp. 335–50 in *Research in the Sociology of Work*, vol. 1, edited by Richard L. Simpson and Ida Harper Simpson. Greenwich, CT: JAI Press.

Teberg, John E. 1986. "Comparing Parameter Estimates Between Linear Models." Pp. 746–52 in SUGI 11: *Proceedings of the Eleventh Annual SAS Users Group International Conference.* Cary, NC: SAS Institute Inc.

U.S. Bureau of the Census. 1973. *Census of Population, 1970. Subject Reports. Occupational Characteristics. Final Report PC(2)-7A.* Washington, DC: Government Printing Office.

Vanneman, Reeve and Fred C. Pampel. 1977. "The American Perception of Class and Status." *American Sociological Review* 42:422–37.

Wright, Erik Olin. 1978. *Class, Crisis, and the State.* New York: New Left Books.

SENTENCING THE WHITE-COLLAR OFFENDER*

MICHAEL L. BENSON   ESTEBAN WALKER

University of Tennessee

With a modified version of the white-collar sentencing model developed by Wheeler, Weisburd, and Bode (1982), we analyze sentences for a sample of white-collar offenders adjudicated in one federal court between 1970 and 1980. Contrary to Wheeler et al., in this study socioeconomic status is not related to sanction severity. Higher-status offenders are no more likely to be incarcerated than low-status offenders, nor do they receive longer sentences. We explore whether the relationship between socioeconomic status and sentence severity changed after the 1974 Watergate scandal. Our analysis suggests that the contextual characteristics of the districts studied by Wheeler et al. may account for their anomalous finding on social status. We argue that contextual variations in criminal sentencing extend both to social status and race.

Sociologists and criminologists long have disagreed whether there are class-based differences in criminal sentencing. Although some investigators report a positive, albeit weak, relationship between class and sentence (Bernstein, Kelly, and Doyle 1977; Chiricos and Waldo 1975), others suggest it is both strong and inverse (Hagan and Parker 1985; Lizotte 1978; Swigert and Farrell 1977). Similar contradictory findings characterize research on race differences in sentencing (Hagan and Bumiller 1983). Against this backdrop of controversy, Wheeler, Weisburd, and Bode (1982), in a major study on the sentencing of white-collar offenders, found that high-status offenders were sentenced *more harshly* than low-status offenders.

In light of this controversial finding, further study of the relationship between status and sentence severity for white-collar offenders is warranted. The present study uses a modified version of the Wheeler et al. sentencing model and analyzes the sentences of white-collar offenders in a different jurisdiction and time period. Although our primary concern is the effect of status on sanction severity, we also explore the impact of shifts in political climate on white-collar sentencing. We conclude with an interpretation of the apparent anomalous finding on social status.

ALTERNATIVE EXPLANATIONS

To explain the effect of socioeconomic status on sentence severity, Wheeler et al. draw upon interviews with judges. They argue that judges are outraged by the crimes of the rich; judges see high-status offenders as particularly blameworthy because of the privileges and advantages that accompany their social and economic positions. Presumably, high-status individuals are bound by a higher level of moral responsibility. Failure to meet this standard makes their crimes more censurable. In addition, Wheeler et al. suggest that a post-Watergate morality may have influenced judges' reactions to white-collar crimes. Sensitized perhaps by criticism of leniency toward white-collar offenders, judges reacted by imposing harsher sentences on them. Because the Wheeler et al. sample only included offenders sentenced between 1976 and 1978, they could not test this interpretation. But this new harshness gains credibility in a recent study of securities violations in Canada (Hagan and Parker 1985). The investigators report a pronounced change in sanctioning of securities violators after "Harbourgate," a major Canadian political scandal. Since our study includes offenders sentenced from 1970 to 1980, we can test whether high-status offenders were sentenced to prison more often and for longer periods after the 1974 Watergate scandal raised public concern about white-collar crimes (Hagan and Palloni 1986).

PROCEDURES

This study was conducted in a federal district court in a midwestern state. The cases

represent individuals sentenced for white-collar crimes between 1970 and 1980.[1] The data were gathered systematically from pre-sentence investigation reports (PSI). Where possible, we followed the procedures used by Wheeler et al. for data collection, reduction, and analysis.

With one exception, the crimes that we designated as white-collar are the same as those used by Wheeler et al. and by Hagan, Nagel-Bernstein, and Albonetti (1980). A new offense, embezzlement or theft of public monies by a public employee, was added. This offense, public embezzlement, consti-tutes 12.6 percent of the sample. Although public embezzlement offenders were not included in the Wheeler et al. study, similar offenses commonly are considered white-collar crime (Conklin 1977; Edelhertz 1970; Coleman 1985). Because it is an abuse of trust, public embezzlement fits the most commonly accepted definition of white-collar crime: a property crime involving deceit or deception (Edelhertz 1970; Shapiro 1980).[2] Also, although the Wheeler et al. sample included securities and exchange, antitrust, and bribery offenders, there were no such offenders in our sample. Our sample contains larger percentages of tax and embezzlement offenders than the Wheeler et al. sample and smaller percentages of lending and credit fraud, false statements, and postal and wire fraud offenders.

Variables

Wheeler et al. developed a complex model of the sentencing process. To enhance compara-bility between our study and theirs, we used a similar model. We divided the independent variables into four categories: act-related; actor-related; legal process; and other. Coding for the legal process and other variables is straightforward and needs no explanation (see

Table 1). The Watergate variable was not included in the Wheeler et al. study.

Based on their interviews, Wheeler et al. argued that judges assess the seriousness of white-collar crimes by applying several crite-ria: the type of victim; the amount of money lost by victims; the complexity and sophisti-cation evidenced in the execution of the offense; and the geographic spread of the offense. Our four act-related variables tap these dimensions of seriousness. Dollar vic-timization measures the amount of money jeopardized or lost by victims. A complexity/sophistication index combines four aspects of the offense: level of organization, patterning, duration, and number of perpetrators. Spread of illegality indicates whether the offense affected only the local community or a wider region. There are four types of victims: individual, business, government, and combi-nation.

The actor-related variables cover offenders' social and criminal backgrounds. Socioeco-nomic status was measured using the proce-dures recommended by Nam and Powers (1983).[3] We used a modified version of the impeccability scale developed by Wheeler et al. This measures the moral and social propriety of a defendant's conduct.[4] White-collar offenders often have contributed posi-tively to the quality of life in their communi-ties, and their moral and social conduct reflects mainstream normative expectations. Thus, they score high on impeccability. Wheeler et al. found that offenders with higher scores on impeccability received more lenient sentences than those with lower scores. In addition to social background, we collected data on defendants' prior criminality (number of prior arrests), role in the offense (minor vs. major), and attitude toward the offense (remorseful vs. not remorseful).

---

[1] The database includes every white-collar case handled in this district between 1970 and 1980 that was available for study. Individuals sentenced after 1980 were not included because we were only permitted access to presentence investigation reports (PSIs) for closed cases. Including post-1980 cases would have limited the study to those receiving the shortest sentences, the ones most likely to be closed. To avoid this truncation bias, 1980 was deemed the cut-off year.

[2] A typical example is an employee of the U.S. Postal Service, who uses his or her position to steal cash or other valuables from the mail.

[3] We used a random-allocation procedure to assign values to missing observations for socioeconomic status and dollar victimization, resulting in a distribution of scores for missing values that approximated the original distribution. Wheeler et al. used the same procedure. The percentages of allocated scores for socioeconomic status and dollar victimization are 8 percent and 12 percent respectively.

[4] We consulted with one of the authors of the Wheeler et al. study on problems encountered with measuring impeccability. Nine of the items used in constructing the original impeccability scale were dropped; the items were either redundant or had too many missing values. The present scale is a composite of 20 items and ranges from a possible low of −11 to a possible high of +17.

Table 1. Description of Variables

	Measurement
Independent Variables	
Act-Related Variables	
Dollar victimization	9-category ordinal scale: range = "0" to "over 2.5 million"
Complexity/sophistication	9-category ordinal scale: range = 4–13
Spread of illegality	3-category ordinal scale: "none" (1), "local" (2), "regional" (3)
Type of victim	4 nominal level variables: "individual," "government," "business," "combination"
Actor-Related Variables	
Socioeconomic status	Interval scale: range = 6.0 – 96.0
Impeccability	Ordinal scale: range = – 11 to 17
Prior arrests	Number of arrests (1, 2, 3, 4+)
Remorse	Binary: remorse (1) vs. no remorse (0)
Role in offense	Binary: minor role (0) vs. alone or major role (1)
Legal Process Variables	
Source of conviction	Binary: plea (0) vs. trial (1)
Nature of counsel	Binary: private (0) vs. public (1)
Statutory offense category	6 nominal-level variables: Tax violations, Postal fraud, False claims, Bank embezzlement, Lending and credit fraud, Public embezzlement
Watergate	Binary: before (1970–1974) (0) vs. after (1975–80) (1)
Other Variables	
Sex	Binary: female (0) vs. male (1)
Age	in years
Race	Binary: nonwhite (0) vs. white (1)
Dependent Variables	
In/out	Binary: imprisonment (1) vs. no imprisonment (0)
Length of sentence	in months (natural log)

There are two dependent variables: incarcerated versus not incarcerated (In/Out) and length of sentence. We used a two-step analytic process. We began by examining the decision whether to incarcerate a defendant or not. Next we examined the length of sentence imposed on those incarcerated.

FINDINGS

The In/Out Decision

We used logistic regression analysis to analyze the In/Out decision. The coefficients in logistic regression can be interpreted as expected changes in the logarithm of the odds ratio, $\ln [p/(1-p)]$, where p is the probability of incarceration. The full model consists of 22 variables, including 3 dummy variables for type of victim and 5 for type of offense (χ^2 = 70.92 with 22 degrees of freedom, p < .0001). The estimated coefficients and their respective p-values are presented in Table 2.

Contrary to Wheeler et al., neither socioeconomic status nor impeccability significantly affects the decision to incarcerate. While our signs for status (positive) and impeccability (negative) are the same as

theirs, the magnitude of the coefficient for socioeconomic status is considerably smaller (about one-tenth) than they reported. At best, social status plays a very small part in the judicial decision whether to incarcerate white-collar offenders. Impeccability has a similar small and insignificant effect on the likelihood of going to prison. The judges for this district apparently are not outraged by the crimes of the rich. Neither are they impressed by the good deeds and exemplary lies of some white-collar offenders. Rather, judges pay little attention either to social status or to social conduct in deciding which defendants merit incarceration.

It is clear in Table 2 that many variables do not contribute significantly to the judge's incarceration decision. We constructed a more parsimonious model using a stepwise procedure. Both forward and backward modes produced similar results. Six variables were retained in the reduced model (χ^2 = 58.92 with 6 degrees of freedom, p < .0001). The variables selected were the ones with the smallest p-values in the full model. The sign and magnitude of the coefficients in the reduced model are similar to the corresponding coefficients in the full model. The statistic

Table 2. Logistic Model of the In/Out Decision, $N = 189$

Variable	Beta	s.e.	Chi-square	p-value
Act-Related				
Dollar victimization	0.328**	0.121	7.31	.0069
Complexity/sophistication	0.176	0.136	1.68	.1954
Spread of illegality	0.954	0.562	2.89	.0894
Type of victim[a]				
Individual	−0.741	0.924	0.64	.4224
Government	−0.500	0.738	0.45	.5014
Combination	−0.280	0.754	0.13	.7145
Actor-Related				
Socioeconomic status	0.002	0.013	0.02	.8787
Impeccability	−0.005	0.073	0.01	.9416
Prior arrests	0.284	0.165	2.97	.0848
Remorse	0.151	0.405	0.14	.7089
Major role in offense	−0.647	0.584	1.23	.2677
Legal Process				
Source of conviction	1.020	0.694	2.16	.1419
Type of attorney	1.205**	0.450	7.16	.0075
Statutory offense[b]				
Tax violations	1.175	0.909	1.67	.1961
Bank embezzlement	0.167	0.764	0.05	.8270
Postal & wire	0.544	0.844	0.42	.5192
False statements	−0.743	1.197	0.39	.5344
Public embezzlement	0.585	1.045	0.31	.5758
Watergate	1.137**	0.435	6.83	.0090
Other				
Sex	0.066	0.475	0.02	.8897
Age	−0.019	0.022	0.75	.3878
Race	−0.727	0.481	2.29	.1303

Intercept = −5.260
Chi-square = 70.92** Degrees of freedom = 22

[a] Reference category is business victim.
[b] Reference category is lending and credit fraud.
* $p < .05$.
** $p < .01$.

for testing the significance of the terms not included in the reduced model is $\chi^2 = 70.92 - 58.92 = 12.00$ with 16 d.f., which has an associated p-value of .256.

After constructing the reduced model, we included socioeconomic status to see if its effect would be more apparent when fewer variables were controlled. The resulting estimated coefficients and p-values for the model appear in Table 3. Despite the smaller

number of independent variables, socioeconomic status is not significant.

In view of the hypothesis that sentencing of white-collar offenders became harsher after Watergate, we wondered if the sentences given the pre-Watergate offenders in the sample were masking the effect of socioeconomic status on post-Watergate sentences. As Table 3 shows, defendants sentenced after 1974 were more likely to be incarcerated than

Table 3. Reduced Logistic Model of the In/Out Decision

Variable	Beta	s.e.	Chi-square	p-value
Socieconomic status	−0.004	0.010	0.16	.6848
Spread of illegality	1.243	0.367	11.44	.0007
Dollar victimization	0.401	0.099	16.57	.0000
Type of attorney	0.935	0.385	5.88	.0153
Watergate	0.914	0.378	5.85	.0156
Race	−0.768	0.390	3.85	.0495
Prior arrests	0.207	0.140	2.19	.1392

Intercept = −4.54
Chi-square = 51.55 ($p < .0001$)
Degrees of freedom = 7

those sentenced before. To test the hypothesis that high-status offenders were sentenced more harshly after Watergate, we estimated the reduced model for offenders sentenced before and after Watergate. In both samples, the coefficient for status was negative and not significant.

As for the other independent variables, the effects of dollar victimization and spread of illegality indicate the importance of the nature of the offense at this stage of the sentencing process. The only significant actor-related variable is the number of prior arrests; prior arrests are directly and positively associated with the probability of incarceration. The positive coefficient of type of attorney shows that the probability of incarceration increases for defendants represented by public attorneys. Contrary to Wheeler et al., we find that nonwhites are more likely to be incarcerated than their white counterparts. The effect of race remained significant even after controlling for socioeconomic status.

Length of Sentence

For defendants who were incarcerated, we employed multiple linear-regression procedures to analyze the relationship between the regressors and the length of sentence.[5] The full model contained 21 variables, including 3 dummy variables for type of victim and 4 for type of offense. Regression diagnostics (Belsley, Kuh, and Welsch 1980) were used to detect influential observations. Two observations were classified as highly influential, which corresponded to individuals who received exceptionally light sentences. Since these observations are not representative of the population, we deleted them from the analysis. We also inspected the data for multicollinearity. The condition number, 42.11, is above the suggested cutoff of 30, but it is mainly associated with the intercept. The largest variance inflation factor (VIF) is 6.19, which indicates that none of the

variable coefficients are adversely affected by multicollinearity (Myers 1986, p. 219). The estimated coefficients and p-values for the full model appear in Table 4.

As with the In/Out decision, social status makes a positive but insignificant contribution to the length of sentence. Since many of the variables did not contribute to the judge's length-of-sentence decision, we again used a stepwise procedure to screen out the noncontributing variables. The results are presented in Table 5.

The coefficient of determination for the model shown in Table 5 is $R^2 = .798$. The contribution of the variables not included in this model is not significant ($R^2 = .823$ for the full model). There were no signs of multicollinearity in the reduced model. Both

Table 4. Multiple Regression Model for Length of Sentence, $N = 77$

Variable	B	T	p-value
Act-Related			
Dollar victimization	0.094	1.640	.1067
Complexity/ sophistication	−0.145*	−2.570	.0129
Spread of illegality	−0.127	−0.683	.4975
Type of victim[a]			
Individual	0.826*	2.325	.0238
Government	−0.094	−0.326	.7454
Combination	0.409	1.647	.1052
Actor-Related			
Socioeconomic status	0.003	0.465	.6437
Impeccability	0.069*	2.119	.0386
Prior arrests	0.029	0.457	.6491
Remorse	−0.338	−1.770	.0822
Major role in offense	0.820**	2.831	.0065
Legal Process			
Source of conviction	0.654*	2.457	.0172
Type of attorney	0.051	0.268	.7896
Statutory offense[b]			
Tax violations	−0.992*	−2.556	.0134
Bank embezzlement	−1.311**	−3.506	.0009
Postal & wire	0.517	1.537	.1300
Public embezzlement	−1.485**	−2.995	.0041
Watergate	−0.017	−0.085	.9328
Other			
Sex	−0.072	−0.336	.7328
Age	0.008	0.863	.3916
Race	−0.511**	−2.811	.0068

Intercept = 2.640
$R^2 = .8230$
$F = 12.177**$

[a] Reference category is business victim.
[b] Reference category is lending and credit fraud. Because only two offenders were incarcerated for false statements and claims, they were dropped from the analysis.
* $p < .05$
** $p < .01$

---

[5] In other analyses, which are available on request, we included a correction term for selection effects in the regression equation for length of sentence. In the full model, this term introduced significant multicollinearities in the data. Multicollinearity was not a problem when the correction term was included in the reduced model. Since the inclusion of the correction term did not significantly affect the magnitudes, signs, or p-values of the variables in the reduced model, we decided not to include it in the analyses reported here.

Table 5. Reduced Regression Model for Length of Sentence

Variable	B	T	p-value
Type of Victim[a]			
Individual	0.829	2.502	.0149
Government	−0.147	−0.539	.5914
Combination	0.271	1.199	.2350
Statutory Offense[b]			
Tax violations	−1.008	−2.991	.0039
Bank embezzlement	−1.691	−5.264	.0001
Postal and wire	0.374	1.174	.2447
Public embezzlement	−1.824	−4.233	.0001
Impeccability	0.093	3.281	.0017
Major role in offense	0.979	3.881	.0002
Source of conviction	0.883	3.916	.0002
Complexity/ sophistication	−0.157	−3.151	.0025
Race	−0.492	−2.892	.0052

Intercept = 3.47
R^2 = .798
F = 21.05 (p < .0001)

[a] Reference category is business victim.
[b] Reference category is lending and credit fraud.

the condition number (18.34) and the largest VIF (5.33) are within the recommended standards for detecting multicollinearity.

That the value for R^2 is larger than usual for a sentencing study probably results from two factors. Because only five judges were involved in sentencing between 1970 and 1980, unmeasured variation among judges is restricted. Additionally, the set of offenses is well defined and seriousness is measured accurately. Hence, extraneous variation within offense type is limited.

For the reduced model, both type of victim and type of offense significantly affect the length of sentence. Five other variables were retained in the model. Since socioeconomic status was not significant even when added to the reduced model, we do not report its results. We also estimated the reduced model with socioeconomic status included for offenders sentenced after Watergate (n = 58). The results mirror those for the full sample.

As with the In/Out decision, race has a significant effect on the length of sentence: Nonwhite defendants received longer sentences than white defendants.

Among the other variables, defendants who played major roles in their offenses received longer sentences than those who played minor roles. As expected, defendants who pled guilty received shorter sentences than those who went to trial. Contrary to our expectations, offenders with high scores on impecca-

bility received longer sentences than those with low scores.

Summary

Our research confirms the importance of the substance of the offense, as measured by dollar victimization and spread of illegality, in predicting incarceration. But there are three major differences between our findings and those of Wheeler et al. While they found that socioeconomic status and impeccability significantly affect the decision to incarcerate, these variables do not predict the probability of incarceration in our sample. We find a significant race effect, and they did not.

Like Wheeler et al., our analysis of the determinants of length of sentence show that actor-related and legal process variables become more important at this stage of the sentencing process than in the In/Out decision. Our results diverge from those of Wheeler et al. at three points: (1) the absence of a significant status effect; (2) the presence of a significant race effect; and (3) the positive effect of impeccability.

EXPLANATION

We began this research with two goals: to replicate the Wheeler et al. finding on social status and to test whether the relationship between social status and sentence severity changed after Watergate. Like Hagan and Palloni (1986), we found that white-collar offenders were incarcerated at a higher rate after 1974, but high-status offenders faced no greater risk of being sent to prison than low-status offenders. Thus, our data do not support the hypothesis that, as part of a broader social movement against white-collar crime (Katz 1980), judges selected out elite offenders for especially severe sentences after Watergate.

That our results contradict those of Wheeler et al. should not be completely surprising. Recall the controversy that surrounds criminal sentencing research. Inconclusive and contradictory results are common and have fueled a debate over the importance of extra-legal factors in the criminal justice system. Sample selection may be one explanation for our results. Our sample included no bribery, antitrust, or securities and exchange offenders. Although the structure of the sample may account for the small status effect, evidence

reported by Wheeler et al. argues against this interpretation. Wheeler et al. (p. 650) examined their data crime by crime and found that the effect of socioeconomic status was stable. In addition, the mean (59.4) and range (9.5 to 95.5) of socioeconomic status in our sample is virtually identical to that reported by Wheeler et al. For these reasons, we believe that sample differences do not account for our divergent findings.

Hoping to explain the inconsistent findings on class and race in sentencing research, some investigators argue that the impact of extra-legal factors varies across social settings (Peterson and Hagan 1984). Increasingly, investigators are examining this contextual variation in criminal sentencing. In light of these recent theoretical developments, we decided to investigate whether there are contextual differences between our district and the Wheeler et al. districts.

One approach to investigating contextual effects focuses on differences in organization and caseload processing among courts. Investigators have found that across judicial districts, caseload and casemix affect the treatment meted out to different types of offenders (Hagan and Bumiller 1983, p. 32). Some evidence shows that large white-collar caseloads give rise to distinctive plea bargaining and sentencing patterns (Hagan and Bernstein 1979; Nagel and Hagan 1982). Consequently, the volume of white-collar prosecutions may affect the severity with which white-collar offenders are sentenced (Hagan et al. 1980).

Wheeler et al. drew their sample from federal district courts located in seven of the largest metropolitan areas in the country. The central cities for the districts are Los Angeles, Atlanta, Chicago, Baltimore, New York, Dallas/Fort Worth, and Seattle. These metropolitan areas were studied because they were known to be the loci of prosecution of many large-scale white-collar crimes (Wheeler et al. 1982, p. 643). Although the Wheeler et al. sampling strategy generated a large and varied collection of white-collar offenses, it limited the sample to cases drawn from districts that are atypical in the number and types of offenses they process.

We used the total number of cases commenced in each district as a measure of caseload and compared the Wheeler et al. districts to all other federal districts in 1975. Six of their districts ranked in the upper

quartile and the seventh was above the median.[6] For two white-collar crimes, fraud and embezzlement, we found that all of the Wheeler et al. districts were in the upper quartile. In fact, four districts ranked in the top eight for these white-collar crimes. Clearly, the districts studied by Wheeler et al. have larger overall caseloads and also handle more white-collar cases than the vast majority of federal district courts. By comparison the district we studied has a more representative caseload of white-collar crimes. The 49 fraud and embezzlement cases commenced there in 1975 place it within one-half a quartile deviation from the median for all federal district courts (Md = 41, QD = 23).

Another approach to specifying contextual effects measures specific features of the social environment. Investigators using this approach have identified urbanization and the proportion of the population that belongs to minority groups (racial mix) as important covariates of criminal sentencing. Several studies document significant rural-urban differences in sentencing of whites and non-whites, with two studies reporting that nonwhites are treated more harshly in rural than in urban settings (Hagan 1977; Pope 1975). Myers and Talarico (1986) found the opposite effect. In their study of juvenile justice, Dannefer and Schutt (1982) report that juvenile courts in areas with large minority populations treat blacks more leniently than courts where minorities make up a smaller percentage of the population.

We compared the Wheeler et al. districts to other federal districts on urbanization and racial mix. As an indicator of urbanization, we used percent of the population residing in urban areas. We found that the Wheeler et al. districts are all above the median (Md = 65 percent), with five ranking in the upper quartile.[7] To measure racial mix, we used percent of the population that is nonwhite and found that six of the seven Wheeler districts

---

[6] We chose 1975 because it falls in the middle of our study period and is close to the time period studied by Wheeler et al. The data on the number of cases commenced was gathered from Table D-3 in the *Annual Report of the Director of the Administrative Office of the U.S. Courts* (Judicial Conference, 1975). Fraud and embezzlement are the only white-collar offenses separated out in the director's report. We excluded Puerto Rico and the three U.S. territories for our calculations.

[7] This information was compiled from the *Federal Judiciary Almanac-1984* (Dornette and Cross 1986).

fall above the median (Md = 14 percent), with four ranking in the upper quartile. By comparison, the district studied here is more representative of federal districts in the percentage of its population that is nonwhite (12 percent) and that resides in urban areas (66 percent).

These data indicate that the districts studied by Wheeler et al. share contextual characteristics that recent sentencing research finds associated with variations in criminal sentencing and processing. Given the urban and racially mixed nature of their districts and their large caseloads, the lack of a race effect in the Wheeler et al. study is not surprising. Significant race effects are less likely in heavily bureaucratized urban courts that process large numbers of cases. These courts are probably too important symbolically and too bureaucratized to permit systematic racial discrimination (Hagan and Bumiller 1983, pp. 31–32). If caseload differences covary with different levels of bureaucratization, then our divergent findings on race are less surprising.

The contextual characteristics of the Wheeler et al. districts may also account for their anomalous finding on social status. If we assume that judges are drawn from the local community and reflect community concerns and values, then the outrage toward high-status white-collar offenders reported among the judges interviewed by Wheeler et al. may be a function of the large urban areas in which these judges live and work. The values of judges in smaller federal districts may differ from those of judges located in major urban centers, where large-scale white-collar crimes happen more often. Moved by concern for the underprivileged, it is largely more liberal urban judges who impose the most severe sentences on high-status white-collar offenders.

Frequent exposure of white-collar crimes threatens the stability of the economic order and the legitimacy of the legal order. In districts where white-collar crimes are common, the legal system is regularly and publicly confronted with challenges to its ideology of equal treatment. To maintain legitimacy, judges sentence harshly the occasional big-time offender unlucky enough to reach the sentencing stage. In districts where white-collar crimes are less frequently exposed, the threat they pose to the legitimacy of the legal order is less. As a result, maintaining legitimacy and achieving general

deterrence through harsh treatment of high-status offenders may be less important for judges in such districts.

It is premature to generalize from the findings in one district, but the absence of a significant status effect after Watergate at either stage of the sentencing process supports our argument that the Wheeler et al. finding on socioeconomic status is due to the contextual characteristics of the districts they sampled. To date, no research has examined the impact of contextual variables, such as urbanization and racial mix, on the sentencing of white-collar offenders. Thus, our study extends the newly emerging research agenda on contextual effects by presenting evidence in an area not previously studied using this approach.

In conclusion, this investigation, like other recent sentencing studies, points to the importance of contextual variations among courts for an enhanced understanding of criminal sentencing. We have shown that the districts studied by Wheeler et al. have larger caseloads, are more urbanized, and are more racially mixed than most federal districts. Our results suggest that these structural and contextual factors, rather than post-Watergate morality, may account for the anomalous finding of Wheeler et al. on social status.

REFERENCES

Benson, Michael L. 1987. White-Collar Sentencing Study [MRDF]. Knoxville: Michael L. Benson, Department of Sociology, University of Tennessee [producer]. Knoxville: Michael L. Benson, Department of Sociology, University of Tennessee [distributor].

Bernstein, Ilene Nagel, William R. Kelly, and Patricia A. Doyle. 1977. "Societal Reaction to Deviants: The Case of Criminal Defendants." American Sociological Review 42:743–55.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Chiricos, Theodore G. and Gordon P. Waldo. 1975. "Socioeconomic Status and Criminal Sentencing: An Empirical Assessment of a Conflict Proposition." American Sociological Review 40:753–72.

Coleman, James W. 1985. The Criminal Elite: The Sociology of White Collar Crime. New York: St. Martin's.

Conklin, John E. 1977. Illegal But Not Criminal: Business Crime in America. Englewood Cliffs, NJ: Prentice-Hall.

Dannefer, Dale and Russell K. Schutt. 1982. "Race and Juvenile Justice Processing in Court and Police Agencies." American Journal of Sociology 87:1113–32.

Dornette, W. Stuart and Robert R. Gross. 1986. Federal Judiciary Almanac-1984. New York: Wiley.

Edelhertz, Herbert. 1970. *The Nature, Impact and Prosecution of White-Collar Crime*. Washington, DC: National Institute of Law Enforcement and Criminal Justice.

Hagan, John. 1977. "Criminal Justice in Rural and Urban Communities: A Study of the Bureaucratization of Justice." *Social Forces* 55:597–612.

Hagan, John and Ilene Nagel Bernstein. 1979. "The Sentence Bargaining of Upperworld and Underworld Crime in Ten Federal District Courts." *Law & Society Review* 13:467–78.

Hagan, John and Kristin Bumiller. 1983. "Making Sense of Sentencing: A Review and Critique of Sentencing Research." Pp. 1–54 in *Research on Sentencing: The Search for Reform*, vol. 2, edited by Alfred Blumstein, Jacqueline Cohen, Susan Martin, and Michael H. Tonry. Washington DC: National Academy Press.

Hagan, John, Ilene Nagel-Bernstein, and Celesta Albonetti. 1980. "The Differential Sentencing of White-Collar Offenders in Ten Federal District Courts." *American Sociological Review* 45:802–20.

Hagan, John and Alberto Palloni. 1986. " 'Club Fed' and the Sentencing of White-Collar Offenders Before and After Watergate." *Criminology* 24:603–22.

Hagan, John and Patricia Parker. 1985. "White-Collar Crime and Punishment: The Class Structure and Legal Sanctioning of Securities Violations." *American Sociological Review* 50:302–16.

Judicial Conference. 1975. Annual Report. Administrative Office of the U.S. Courts, 1975. Washington, DC: U.S. Government Printing Office.

Katz, Jack. 1980. "The Social Movement Against White-Collar Crime," Pp. 161–84 in *Criminology*

Review Yearbook, vol. 2, edited by Egon Bittner and Sheldon Messinger. Beverly Hills, CA: Sage.

Lizotte, Alan J. 1978. "Extra-legal Factors in Chicago's Criminal Courts: Testing the Conflict Model of Criminal Justice." *Social Problems* 25:564–80.

Myers, Martha A. and Susette M. Talarico. 1986. "Urban Justice, Rural Injustice? Urbanization and its Effect on Sentencing." *Criminology* 24:367–92.

Myers, Raymond H. 1986. *Classical and Modern Regression with Applications*. Boston: Duxbury.

Nagel, Ilene H. and John Hagan. 1982. "The Sentencing of White-Collar Criminals in Federal Courts: A Socio-Legal Exploration of Disparity." *Michigan Law Review* 80:1427–65.

Nam, Charles B. and Mary G. Powers. 1983. *The Socioeconomic Approach to Status Measurement*. Houston: Cap and Gown Press.

Peterson, Ruth D. and John Hagan. 1984. "Changing Conceptions of Race: Towards an Account of Anomalous Findings of Sentencing Research." *American Sociological Review* 49:56–70.

Pope, Carl E. 1975. *Sentencing of California Felony Offenders*. National Criminal Justice Information and Statistics Service. Washington, DC: U.S. Department of Justice.

Shapiro, Susan. 1980. *Detecting Illegalities: A Perspective on the Control of Securities Violations*. Ph.D. diss., Yale University.

Swigert, Victoria Lynn and Ronald A. Farrell. 1977. "Normal Homicides and the Law." *American Sociological Review* 42:16–32.

Wheeler, Stanton, David Weisburd, and Nancy Bode. 1982. "Sentencing the White-Collar Offender: Rhetoric and Reality." *American Sociological Review* 47:641–59.

COMMENTS

◆

CRIME, DETERRENCE, AND CHOICE: TESTING THE RATIONAL BEHAVIOR HYPOTHESIS*

JOHN M. HEINEKE
Santa Clara University

In their recent paper, Piliavin, Gartner, Thornton, and Matsueda (1986) (PGTM hereafter) discuss how best to measure the deterrent effects of criminal sanctions—if such exist. Of particular value is their attempt to integrate the considerably different approaches taken by economists and sociologists. However, as an economist whose work in two areas was referenced by PGTM, I would like to make several comments on the PGTM paper: (1) comments concerning methodological differences at the empirical level; (2) comments that are statistical in nature and concern the validity of the PGTM empirical results (which consequently call into question their finding that sanction levels have little impact on offender decision); and (3) most important, I would like to offer some observations on the authors' claim that the paper's empirical exercise constitutes a test of the "rational-choice" model. It is important to emphasize at the outset that the purpose of my remarks is to stimulate cross-disciplinary discussion—a potentially fertile source of new ideas.

1. The authors correctly observe that, in general, measures of rewards and sanctions in a model of rational choice must be measures of *perceived* rewards and *perceived* sanctions, not measures of *actual* rewards and *actual* sanctions, if tests of the model are to be valid. But there is an important case where actual rewards and sanctions may be used—instead of their perceived counterparts—when tests of a model are qualitative (i.e., one is interested in signs of coefficients and not in magnitudes of coefficients), as in PGTM. In this situation one can use actual reward and sanction levels from the public record *if* the latter are

increasing monotonic functions of the former. Symbolically, let x^a and x^p represent actual and perceived rewards or sanctions. Then if $x^p = f(x^a)$ and if f is a monotonically increasing function, then signs of estimated coefficients remain invariant and tests using actual rewards and sanctions yield the same *qualitative* results as tests using perceived rewards and sanctions. Notice that all this condition requires is that when rewards and sanctions change, criminals or potential criminals *do not* conclude that the change is in the opposite direction. Of course whether or not perceptions about sanctions(rewards) are monotonically related to actual sanction(rewards) levels is an empirical question, but it seems likely that this condition will often be met; if so, it frees the researcher from using what may be highly unreliable self-reports of perceived rewards and sanctions. This point is expanded upon in the next paragraph.

PGTM note that economists seem to prefer macro-level analyses of criminal behavior. They are correct, although *ceteris paribus,* economists would much prefer to use micro-level data and fit micro-models of human behavior. So why all the macro models? One important reason is a basic distrust of many interview- or questionnaire-generated data sets. Between the Scylla of macro models and the Charybdis of micro models, in which the sample is generated from surveys, questionnaires, or interviews with less-than-disinterested parties, economists often choose to estimate macro-level models. The literature calling into question the validity of such survey/questionnaire/interview-based data sets is very large. Much of this literature originates in psychology or applied psychology areas and appears in such diverse disciplines as marketing and law(witness veracity). To get a quick sense of the flavor of this literature, consider the Wind and Lerner (1979) paper, which compares the results of consumer interviews designed to elicit information about past purchases with *actual* past purchases (taken from daily purchase diaries). The results are startling: 15 percent of the respondents reported not ever having tried a product when, in fact, it is recorded in their purchase records: 48% who designated a

* Direct all correspondence to John M. Heineke, Department of Economics, Santa Clara University, Santa Clara, CA 95053.

given brand as their "most often bought brand" had *never* purchased the brand. Wind and Lerner conclude that the inaccuracy of such historical purchase data presents a major problem for users of interview-based purchase data. This theme is echoed repeatedly throughout this literature (see also Sudman and Ferber 1974, Weitz and Wright 1979, and Warshaw 1980). In the case at hand, the problem would appear to be even worse. Here convicted felons were asked to recall details of their criminal careers. The veracity of such interview responses would appear to be highly suspect. What incentive is there for a convicted felon to provide an accurate response? If the literature referenced above is even remotely accurate, it is likely that braggarts in these groups will inflate variables such as crimes committed or street income and understate what PGTM call formal-risk variables—*if* they can even remember the details. On the other hand, group members who do not trust the interviewer may well underreport crimes committed or street income and overstate their perceptions of formal risk. The net result of such a diversity of response incentives on the part of interviewees quickly leads to a biased sample and unwarranted inferences. As a result, economists often use publicly reported (objective) macro-data sets, preferring to confront the attendant aggregation and identification problems rather than confront bias problems stemming from sampling errors that arise from interview-based data sets.

2. I now turn to the question of the validity of the inferences made in PGTM. In particular, why were not the sanction variables statistically significant? One reason could be the reporting bias discussed above, but there are other problems as well. PGTM have assumed that their independent variables are distributed following the multivariate normal distribution. As they hint in note 9, dichotomous and ordinal variables cannot be normally distributed (although the majority of their variables are of this type). Since factors are assumed to be linearly related to observable variables, it follows that dependent variables *must* be normally distributed in their model. But as long as these variables are dichotomous or ordinal, they cannot be normal or even approximately normal. In this case, variances of factors will not be constant over the sample, and tests of coefficient significance

and overall goodness of fit based (ultimately) on the normal distribution will be invalid.

But this is not as serious as the problem that arises when linear functions are estimated using dependent variables that (1) cannot be negative, and (2) often take on the value zero, as in PGMT. In this case, the absence of negative values for dependent variables tends to push the estimated linear relationship above the axis. This amounts to forcing the estimated relationship between dependent and independent variables to be flat, and hence implies that coefficients are underestimated! In other words, dichotomous and/or ordinal dependent variables based at zero in a linear relation lead to coefficient estimates that are *biased toward zero* and this bias persists *for all sample sizes*, i.e., parameter estimates are inconsistent. (See Judge, Griffiths et al. (1980) or Goldberger (1964) for more detail on these problems.) This point is fundamental since no amount of data will prevent coefficients from being strongly biased toward zero.[1] And this is precisely the sanction significance question: Are sanction coefficient estimates significantly different from zero? Clearly one cannot answer this question in this statistical framework. It would appear that existence of this latter problem, especially when coupled with the other points mentioned above, goes a long way toward explaining the poor showing of goodness of fit statistics and the general lack of significance of sanction statistics.

3. Finally I would like to comment on PGTM's claim to have provided a test of the "rational choice" model. The basis for this conclusion is the fact that they have included a number of reward and sanction variables in their statistical model along with other variables. Unfortunately, including several relevant variables into a statistical model provides but a very weak test of the rational choice model. A more powerful test of whether actors behave in the crime arena *as if* they were making rational choices would use many more of the implications of that model. One of the most powerful means of accomplishing this is to call upon the theorems of duality theory (see e.g., Diewert (1974) for an overview), which permit the investigator

---

[1] Notice that this line of reasoning implies that even parameter estimates that were found to be statistically significant by PGTM will be biased toward zero—that is, underestimated in magnitude.

to embed into the structure of his/her statistical model all of the theoretical implications of the rational choice model. The point here is that so-called "rational behavior" means that actors behave as if they were maximizing their own expected utility. This behavioral hypothesis places many theoretical restrictions on the behavior of the actor. These restrictions show up in the equation system that comes about when the expected utility function is maximized. The basic procedure in using duality theory is to write down a general expected utility function *and* the set of equations forthcoming when this function is maximized. This entire set of equations is then estimated simultaneously with all of the restrictions on individual equations that flow from expected utility maximization. Likelihood-ratio tests are used to contrast the equation system generated under the assumption of rational behavior with the analogous equation system that is unencumbered by the restrictions implied by the rational-behavior hypothesis.

This procedure provides a powerful test of the rational-behavior hypothesis. Interested readers can find applications of duality theory in a criminal-behavior context in Heineke (1978) and Darrough and Heineke (1978). The main point here is that one can write down numerous models containing reward and sanction variables that are inconsistent with rational behavior.

Finally it should be noted that there are many studies based on both micro and macro models that find statistically significant deterrent effects. I concur with the implicit conclusion of PGTM—for many of the reasons they give—that the majority of statistical studies of deterrence in economics, psychology, and sociology are badly flawed and of little value. But there are many studies where aggregation, identification, and model specification issues have been adequately addressed, and strong deterrent effects have been found. (See e.g., Ehrlich (1973, 1975), Votey (1979, 1982)).

REFERENCES

Darrough, M.N. and J.M. Heineke. 1978. "The Multi-Output Translog Production Cost Function: The Case of Law Enforcement Agencies." Pp. 259–302 in *Economic Models of Criminal Behavior*, edited by J.M. Heineke. Amsterdam: North-Holland.

Diewert, W.E. 1974. "Applications of Duality Theory." Pp. 106–99 in *Frontiers in Quantitative Economics*, vol. II, edited by M.D. Intriligator and D.A. Kendrick. Amsterdam: North-Holland.

Ehrlich, I. 1973. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy* 81:521–65.

———. 1975. "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review* 65:397–417.

Judge, G., W.E. Griffiths, R. Carter Hill, H. Lutkepohl, and T. Lee. 1980. *The Theory and Practice of Econometrics*. New York: John Wiley and Sons.

Goldberger, A. 1964. *Econometric Theory*. New York: John Wiley & Sons Inc.

Heineke, J.M. 1978. "Substitution Among Crimes and the Question of Deterrence: An Indirect Utility Function Approach to the Supply of Legal and Illegal Activity." Pp. 153–210 in *Economic Models of Criminal Behavior*, edited by J.M. Heineke. Amsterdam: North-Holland.

Sudman, S. and R. Ferber. 1974. "A Comparison of Alternative Procedures for Collecting Consumer Expenditure Data for Frequently Purchased Products." *Journal of Marketing Research* 11:128–35.

Piliavin, I., Rosemary Gartner, Charles Thornton, and Ross L. Matsueda. 1986. "Crime Deterrence and Choice." *American Sociological Review* 51:101–19.

Votey, H.L. 1979. "Detention of Heroin Addicts, Job Opportunities and Deterrence." *Journal of Legal Studies* 8:585–606.

———. 1982. "Scandinavian Drinking-Driving Control: Myth or Reality." *Journal of Legal Studies* 11:93–116.

Warshaw, P. 1980. "Predicting Purchase and other Behaviors from General and Contextually Specific Intentions." *Journal of Marketing Research* 17:26–33.

Weitz, B. and P. Wright. 1979. "Retrospective Insight on Factors Considered in Product Evaluation." *Journal of Consumer Research* 6:280–94.

Wind, Y. and D. Lerner. 1979. "On the Measurement of Survey Data: Surveys Versus Purchase Diaries." *Journal of Marketing Research* 16:39–47.

ECONOMIC ASSUMPTIONS VERSUS EMPIRICAL RESEARCH*

(Reply to Heineke, *ASR,* this issue)

ROSS L. MATSUEDA   IRVING PILIAVIN
University of Wisconsin-Madison

ROSEMARY GARTNER
University of Iowa

Heineke raises three major criticisms of our paper. First, he maintains that self-reported

measures of crime are invalid compared to aggregate official statistics from the criminal-justice system. Second, he argues that our statistical models may be invalid because some of our variables are not normally distributed, are not continuous, and are truncated at zero. Third, he suggests that because our model does not incorporate additional overidentifying restrictions, it constitutes a weak test of rational choice. Most of these criticisms are addressed directly in our paper; others are speculative defenses of assumptions found in aggregate-level analyses of individual-level models of criminal decision-making; still others are better termed "proposals for new research." Thus, while we welcome critical attention to our research and are pleased to have stimulated cross-disciplinary discourse, we view Heineke's remarks more as a defense of his own research relying on aggregate official statistics than as a contribution to our line of research.

Let us state our position on aggregate analyses of official statistics: We are skeptical of the inferences made about individual-level processes using aggregate data, but we do not, in principle, dismiss them out of hand. Rather, our position is that the current state of aggregate-level models has not adequately solved the problems of aggregation, measurement error, and identification. Therefore, individual-level models, which measure theoretical concepts directly and at the correct level of analysis, can shed important light on criminal decision-making. Heineke raises several important questions concerning individual-level analyses of survey data, but presents selective evidence to illustrate his arguments. Our reply seeks to redress this imbalance.

Heineke's first criticism concerns the relative validity and reliability of self-reports from surveys versus official records from criminal-justice agencies. While he quickly dismisses self-reports out of hand, he ignores much of the pertinent empirical research, choosing instead to speculate about possible sources of invalidity and cite studies of questionable relevance. Presenting evidence from marketing studies of consumer behavior, Heineke notes that when consumers are asked to report which brands of margarine, soft drinks and birth control devices they recently purchased, their answers fail to correspond to information gleaned from daily purchase

diaries. Based on this evidence, Heineke concludes that all "survey/questionnaire/interview" data are drawn into question. Such results must begin with the assumption that diaries yield more valid data that do surveys. But more importantly, of what relevance to recalling recent criminal behavior is forgetting as trivial an event as purchasing one brand of margarine, soft drink, or birth control device over another? Heineke ignores the substantial literature addressing the relative validity and reliability of self-reported crime versus official statistics—*a literature that is cited in our paper* (Piliavin, Gartner, Thornton, and Matsueda 1986, p. 108). Far from dismissing the self-report method out of hand, the best research on its validity and reliability finds it comparable to most social science measures, and probably superior to official records for the purpose of testing theories (Hindelang, Hirschi, and Weis 1979, 1981; Elliott and Ageton 1980; Huizinga and Elliott 1984, 1986). At the same time, however, these studies caution researchers in using self-reports indiscriminately: for some purposes, such as estimating the true national crime rate, and for some populations, such as black males, the method is less valid. Nevertheless, none of these studies rejects the use of self-reported crime.

Heineke also questions our measures of perceived risks of formal sanction and perceived likelihood of returns to crime. While agreeing with us that rational choice pertains to rewards and costs of criminal and conventional alternatives *as perceived by the actor,* Heineke nevertheless rejects direct (self-report) measures of perceived rewards and costs in favor of indirect (official crime statistics) measures. Referring to official statistics as "actual rewards and punishments," he argues that they are superior to self-reports *if we can assume that "actual" rewards and sanctions are monotonic functions of their perceived counterparts.* Specifically, he maintains that the parameter estimates of a model using aggregate rates of punishment may differ from the parameters of a true population model of perceived risks, but the signs of coefficients will be correct.

The strategy advocated by Heineke simply assumes away what is, in fact, an empirical issue. There is little empirical evidence that aggregate measures of rewards and sanctions, constructed from police, court, and prison records, are strictly monotonically increasing

functions of individuals' perceptions of rewards and sanctions of crime (see Nagin 1978; Zimring and Hawkins 1973). This issue, concerning the relative validity and reliability of the alternate measures, is conceptualized in Figure 1. Note that Heineke's strategy requires two assumptions: (1) true subjective perceptions of the likelihood of costs and returns are monotonically related to the true (aggregate) likelihood of objective risks and returns; and (2) true objective risks and returns are monotonically related to fallible measures of risks and returns computed from aggregated criminal justice reports. That is, in Figure 1, both λ_x and β must reflect monotonic relationships. Some researchers would question whether true perceived certainty is monotonically related to true objective certainty; however, the empirical evidence is sparse, indirect, and unclear. Some quasi-experimental research finds that changes in criminal-justice practices, when highly publicized, alter criminal behavior temporarily (Ross 1973; Press 1971). Other research, however, finds no effect (Kelling and Pate 1974; Joint Committee on New York Drug Law Evaluation 1978; Robertson, Rich, and Ross 1973).

Even more researchers would question whether true objective certainty of punishment is monotonically related to certainty measured by official statistics (see Nagin 1978; Zimring and Hawkins 1973). Aggregate measures of certainty of punishment are constructed by dividing rates of punishment (such as arrest, conviction, or incarceration) by a measure of the crime rate (such as crimes known to the police or arrests). A number of studies have examined the process of producing criminal-justice records, including both the numerator and denominator of indexes of certainty of punishment. These studies suggest that such records are not monotonically related to their true counterparts, but instead reflect, in part, organizational exigencies (Kitsuse and Cicourel 1963; Cicourel 1976; Wilson 1968) and political processes (Skolnick 1966; Komesar 1972; Seidman and Couzens 1974; Campbell 1969). Moreover, even if we could assume that these measures are reasonably valid and reliable, their use in an index of certainty of punishment may still be flawed. Because such indexes must include some measure of criminal behavior in the denominator, they cannot distinguish changes in true certainty of punishment from subsequent changes in criminals' attempts to avoid detection by being more careful (Cook 1980).

Heineke's strategy has other problems. He is incorrect in claiming that the monotonic assumption requires only that "when rewards and sanctions change, criminals or potential criminals *do not* conclude that the change is in the opposite direction." Clearly, if true perceived certainty of punishment is relatively insensitive to subtle changes in the arrest rate, results based on official statistics will be biased. Furthermore, Heineke is incorrect in saying that in our study, we were only interested in signs of coefficients, not magnitudes. In fact, we were acutely interested in the relative magnitudes of coefficients in our model. Finally, and perhaps more significantly, Heineke's argument that the signs of coefficients will necessarily be correct given the monotonic assumption is specious; counter examples are not difficult to construct.

Heineke's second criticism concerns our finding of a null effect of perceived sanctions on criminality. He argues that the null finding is due to departures from multivariate normality, which renders significance tests invalid. In fact, we considered this possibility, and discuss it in the paper (Piliavin et al. 1986, pp. 107–15). Heineke, however, does not mention that we estimated logistic regression models and failed to find significant effects for our sanction variables (Piliavin et al. 1986, p. 107). Nor does he acknowledge, as we did (Piliavin et al. 1986, p. 107), that Monte Carlo studies of the maximum-likelihood estimator of the LISREL program
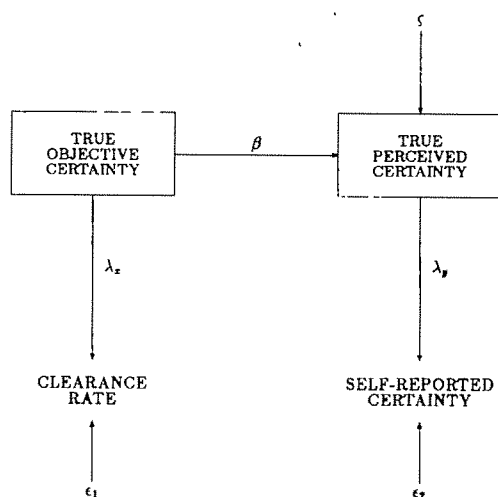
Fig. 1. Diagram of the Relationship Between Alternate Measures of the Certainty of Punishment

find relatively robust to modest departures from normality when the sample is large (Boomsma 1982, 1983). In a similar vein, Heineke surmises that the poor showing of our sanction variables could be due to floor effects in our measures. Once again, we considered this possibility and discussed it in our paper (Piliavin et al. 1986, pp. 114–15). Indeed we tried to squeeze out a significant deterrent effect by examining several non-linear functional forms; none of them succeeded. In short, Heineke has offered several plausible scenarios, all of which we considered in our paper, and none of which is supported in the data.

In his third criticism, Heineke argues that our model constitutes a weak test of rational choice. Although we included the relevant variables in our model, he maintains that we could have employed a more powerful test by incorporating additional overidentifying restrictions implied by a utilitarian model. As an exemplary model, Heineke cites his own work (Heineke 1978). While we agree that other rational-choice models need to be investigated and that our model is far from the final word on rational choice, we nevertheless have difficulties with Heineke's proposals. For example, in the "exemplary" work, Heineke derives from utility-maximization theory a system of legal work supply functions (Model H_0), which he treats as part of his maintained hypothesis. He then specifies two alternative models (Models H_A and H_B), which incorporate additional behavioral constraints consistent with his version of rational choice. Thus, Models H_A and H_B, being more restrictive and nested within the Model H_0, can be tested against Model H_0 using the likelihood-ratio method. Applying such a test, Heineke rejects Model H_A and H_B, since they fail to fit better than Model H_0. But because Model H_0 was part of his maintained hypothesis, Heineke never tests it; its truth—and the truth of its utilitarian assumptions and deterrent effects—is simply assumed.[1]

Finally, Heineke cites the works of Ehrlich (1973, 1975) as exemplars of macro models

that find deterrent effects while adequately addressing problems of specification, identification, aggregation, and measurement. This hits at the heart of our disagreement with Heineke; we find Ehrlich's treatment of these issues unconvincing. We agree with Ehrlich's critics—economists, statisticians, and sociologists—who argue that such models are probably misspecified, underidentified, and plagued with measurement errors (Manski 1978; Fisher and Nagin 1978; Brier and Fienberg 1980).

In conclusion, Heineke has offered several speculative defenses of his own work and that of other economists who work with aggregate data. He has not, however, offered any strong evidence to support his speculation or to draw our results into serious question. What is needed is not speculation but research. Thus, we would welcome further research investigating the assumption that objective measures of sanctions are monotonically related to subjective sanctions; investigating the robustness of the LISREL estimation procedure for specific models such as ours; and investigating the issue of deterrence from alternative rational-choice models of crime. If our research succeeds in stimulating one or more of these lines of inquiry, we will have fulfilled one of our major goals.

REFERENCES

Brier, Stephen S. and Stephen E. Fienberg. 1980. "Recent Econometric Modeling of Crime and Punishment." *Evaluation Review* 4:147–91

Boomsma, Ann. 1982. "The Robustness of LISREL Against Small Sample Sizes in Factor Analysis Models." Pp. 149–73 in *Systems Under Indirect Observation: Causality, Structure, Prediction*, edited by K.G. Joreskog and H. Wold. Amsterdam: North Holland.

———. 1983. *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Non-Normality*. Ph.D. diss. University of Gronigen, The Netherlands.

Campbell, Donald. 1969. "Reforms as Experiments" *American Psychologist* 24:409–29.

Cicourel, Aaron V. 1976. *The Social Organization of Juvenile Justice*. London: Heinemann.

Cook, Phillip J. 1980. "Research in Criminal Deterrence: Laying the Groundwork for the Second Decade." Pp. 211–68 in *Crime and Justice: An Annual Review of Research*, edited by Norval Morris and Michael Tonry. Chicago: University of Chicago Press.

Ehrlich, Issac. 1973. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy* 81:521–65.

———. 1975. "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review* 65:397–417.

---

[1] Our model of costs and returns corresponds to Heineke's maintained hypothesis, which we tested empirically. We found empirical support for the return component and no support for the cost component. Since Heineke's proposed alternative models are more restrictive than our rejected model, they could not possibly fit our data.

Elliott, Delbert S. and Suzanne S. Ageton. 1980. "Reconciling Race and Class Differences in Self-Reported and Official Estimates of Delinquency." *American Sociological Review* 45:95–110.

Fisher, Franklin M. and Daniel Nagin. 1978. "On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rates and Sanction Levels." Pp. 361–99 in *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, edited by Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin. Washington, DC: National Academy of Sciences.

Heineke, John M. 1978. "Substitution Among Crimes and the Question of Deterrence: An Indirect Utility Function Approach to the Supply of Legal and Illegal Activity." Pp. 153–210 in *Economic Models of Criminal Behavior*, edited by John M. Heineke. Amsterdam: North-Holland.

Hindelang, Michael J., Travis Hirschi, and Joseph G. Weis. 1979. "Correlates of Delinquency: The Illusion of a Discrepancy." *American Sociological Review* 44:995–1014.

————. 1981. *Measuring Delinquency*. Beverly Hills: Sage.

Huizinga, David and Delbert S. Elliott. 1984. "Self-Reported Measures of Delinquency and Crime: Methodological Issues and Comparative Findings." Unpublished ms. Boulder, CO: Behavioral Research Institute.

————. 1986. "Reassessing the Reliability and Validity of Self-Report Delinquency Measures." *Journal of Quantitative Criminology* 2:293–327.

Joint Committee on New York Drug Law Enforcement. 1978. *The Nation's Toughest Drug Law: Evaluating the New York Experience*. Washington, DC: U.S. Government Printing Office.

Kelling, George L. and Anthony Pate. 1974. *The Kansas City Preventive Patrol Experiment*. Washington, DC: The Police Foundation.

Kitsuse, John I. and Aaron V. Cicourel. 1963. *The Educational Decision-Makers*. Indianapolis: Bobbs-Merrill.

Komesar, Yale. 1972. "How to Use, Abuse and Fight Back with Crime Statistics." *Oklahoma Law Review* 25:239–58.

Manski, Charles F. 1978. "Prospects for Inference on Deterrence Through Empirical Analysis of Individual Criminal Behavior." Pp. 400–24 in *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, edited by Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin. Washington, DC: National Academy of Sciences.

Nagin, Daniel. 1978. "General Deterrence: A Review of the Empirical Evidence." Pp. 95–139 in *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*, edited by Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin. Washington, DC: National Academy of Sciences.

Piliavin, Irving, Rosemary Gartner, Charles Thornton, and Ross L. Matsueda. 1986. "Crime, Deterrence, and Rational Choice." *American Sociological Review* 51:101–19.

Press, S. James. 1971. *Some Effects of an Increase in Police Manpower in the 20th Precinct of New York City*. New York: Rand Institute.

Robertson, Leon, Robert F. Rich, and H. Laurence Ross. 1973. "Jail Sentences for Driving While Intoxicated in Chicago: A Judicial Policy That Failed." *Law and Society Review* 8:55–68.

Ross, H. Laurence. 1973. "Law, Science, and Accidents: The British Road Safety Act of 1967." *Journal of Legal Studies* 2:1–78.

Seidman, David and Michael Couzens. 1974. "Getting the Crime Rate Down: Political Pressure and Crime Reporting." *Law and Society Review* 8:457–94.

Skolnick, Jerome H. 1966. *Justice Without Trial: Law Enforcement in Democratic Society*. New York: Wiley.

Wilson, James Q. 1968. *Varieties of Police Behavior*. New York: Altheneum.

Zimring, Franklin and Gordon Hawkins. 1973. *Deterrence: The Legal Treat in Crime Control*. Chicago: University of Chicago Press.

ASA Journals

Current and Forthcoming Articles

Social Psychology Quarterly (formerly Sociometry)
Vol. 50, No. 4 (December 1987)

Can Functions be Measured? A New Perspective on the Functional Approach to Attitudes, *Gregory M. Herek*

The Impact of Conjoint Threats to Social- and Self-Esteem on Evaluation Apprehension, *Mark R. Leary, Byron D. Barnes, Carol Griebel, Eileen Mason, and Don McCormack, Jr.*

Children and Civility: Ceremonial Deviance and the Acquisition of Ritual Competence, *Spencer E. Cahill*

Norms, Preferences, Identities, and Retention Decisions, *Bruce J. Biddle, Barbara J. Bank, and Ricky L. Slavings*

Information Patterns, Attribution, and Attraction, *Charles E. Kimble and Frank R. Kardes*

A Confirmatory Factor-Analytic Model of Alienation, *Bruce R. Roberts*

Corrigendum: Methodology, Technology, and Serendipity, *Robert J. Johnson and Howard B. Kaplan*

Journal of Health and Social Behavior
Vol. 28, No. 4 (December 1987)

Health and the Life Course: Some Personal Observations, *John A. Clausen*

The Relative Contributions of Endocrine Changes and Social Circumstances to Depression in Mid-Aged Women, *John B. McKinlay, Sonja M. McKinlay, and Donald Brambilla*

Adolescent Competence, Psychological Well-Being, and Self-Assessed Physical Health, *David Mechanic and Stephen Hansell*

Physicians' Communication Style and Patient Satisfaction, *Mary Klein Buller and David B. Buller*

Interpreting Physician Participation in Randomized Clinical Trials: The Physician Orientation Profile, *Kathryn M. Taylor and Merrijoy Kelner*

HMO Satisfaction and Understanding Among Recent Medicare Enrollees, *Russell A. Ward*

Contemporary Sociology
Vol. 17, No. 1 (January 1988)

Patricia Kaplan, *Class and Gender in India: Women and Their Organizations in a South Indian City*. Lina Fruzzetti and Marilyn Rueschemeyer

Susan Dorsky, *Women of Amran: A Middle Eastern Ethnographic Study*. Lina Fruzzetti and Marilyn Rueschemeyer

Scarlett T. Epstein, Kate Crehan, Annemarie Gerzer, and Jurgen Sass, *Women, Work, and Family in Britain and Germany*. Lina Fruzzetti and Marilyn Rueschemeyer

Joanna Liddle and Rama Joshi, *Daughters of Independence: Gender, Caste, and Class in India*. Lina Fruzzetti and Marilyn Rueschemeyer

Elizabeth Vallance and Elizabeth Davies, *Women of Europe: Women MEPs and Equality Policy*. Lina Fruzzetti and Marilyn Rueschemeyer

Michael Goldfield, *The Decline of Organized Labor in the United States*. William Form

John A. Hall, *Powers and Liberties: The Causes and Consequences of the Rise of the West*. Jack A. Goldstone

Evelyne Huber Stephens and John D. Stephens, *Democratic Socialism in Jamaica: The Political Movement and Social Transformation in Dependent Capitalism*. Walter Korpi

J. A. Mangan, *The Games Ethic and Imperialism: Aspects of the Diffusion of an Ideal*. Gary Alan Fine

Michael Rosenthal, *The Character Factory: Baden Powell's Boy Scouts and the Imperatives of Empire*. Gary Alan Fine

Teaching Sociology
Vol. 16, No. 1 (January 1988 Special Issue on Teaching Complex Organizations)

Teaching the Sociology of Complex Organizations: Issues and Strategies, *Thomas E. Drabek*

Teaching Sociology of Complex Organizations: An Experiential Approach, *Esther Ngan-Ling Chow and William E. Hemple*

Teaching Complex Organizations: A Twenty-Year Odyssey, *Douglas S. Snyder*

A Case-Study Method for Teaching Industrial Sociology, *Vern Baxter*

Taking the Role of the Other: An Aid to Marketing Applied Sociology, *Gene Pertrus and Raymond J. Adamek*

A Capstone Course in Applied Sociology, *Richard Cheever Wallace*

Undergraduate Education in Sociology as Career Training: The Case for an Internship Program, *M. Herbert Danzger*

Micro-Computer Test-Generation Systems: A Software Review, *Victoria Sturtevant and Bryce Johnson*

DEMOS: Teaching the Demography Unit Through Computer Simulations, *Steven Stack*

Also four special notes

I

II

For your sociology bookshelf from Chicago '88

III

V

New Books on Social Theory
from
Greenwood Press and Praeger Publishers

THE END OF CONVERSATION
The Impact of Mass Media on Modern Society
by Franco Ferrarotti
Ferrarotti offers a provocative look at the future of a world dominated by mass media—particularly television. He argues that the art of story-telling, the traditions of oral history, the simple pleasures of individual conversation are being lost because they embody qualities antithetical to the technological imperatives of mass society.
Available May 1988. ISBN 0-313-26087-7. $33.00 est. Greenwood Press

DRAMATURGICAL ANALYSIS OF SOCIAL INTERACTION
by A. Paul Hare and Herbert H. Blumberg with selections from the works of Erving Goffman, David A. Snow, Louis A. Zurcher and Robert Peters, R.S. Perinbanayagam, Ronny Turner, and Charles Edgley
This text provides a summary and examples of generally accepted views of dramaturgical analysis. Beginning with an introduction to and definition of the key concepts of dramaturgical analysis, it goes on to discuss and apply this theory to the analysis of social interaction.
ISBN 0-275-92762-8. $38.95 Praeger Publishers

NEW PERSPECTIVES ON SOCIAL CLASS AND SOCIOECONOMIC DEVELOPMENT IN THE PERIPHERY
edited by Nelson W. Keith and Novella Zett Keith
This volume focuses on the economic features that make Third World social formations distinctive and on nonproperty characteristics such as religion, ethnicity, and culture, that are central to the survival of these societies. The contributors look at the significance and revolutionary potential of peasant majorities, who take limited advantage of capitalist modes of production and often manage to maintain their cultural and economic identity and a degree of independence in the process.
ISBN 0-313-25688-8. $38.00 est. Greenwood Press

THE DEVELOPMENT OF A POSTMODERN SELF
A Computer-Assisted Comparative Analysis of Personal Documents
by Michael R. Wood and Louis A. Zurcher
This study collectively describes the postmodern self—changes in identity and culture that will accompany the "postindustrial revolution" forecast for the end of the twentieth century. It provides a framework for understanding the ideas that collectively comprise postmodern self theory and presents empirical data concerning its historical validity.
ISBN 0-313-25458-3. $37.95 Greenwood Press

FAMILY AND THE
FEMALE LIFE COURSE
The Women of Verviers,
Belgium, 1849–1880
George Alter
$39.50 Cloth / $24.25 Paper, May

BELONGING IN AMERICA
Reading between the Lines
Constance Perin
$24.50 Cloth, April

SOCIALIST ENTREPRENEURS
Embourgeoisement in
Rural Hungary
Ivan Szelenyi
$35.00 Cloth / $15.95 Paper

THE END OF
ORGANIZED CAPITALISM
Scott Lash and John Urry
$45.00 Cloth / $15.00 Paper

THE POLITICAL ECONOMY
OF INFORMATION
*Vincent Mosco and
Janet Wasko, editors*
$40.00 Cloth / $17.50 Paper, April

TECHNOLOGIES
OF CONTROL
The New Interactive
Media for the Home
Kevin G. Wilson
$37.50 Cloth / $15.25 Paper, June

GENTRIFICATION AND
DISTRESSED CITIES
An Assessment of Trends in
Intrametropolitan Migration
Kathryn P. Nelson
$39.50 Cloth / $16.25 Paper, April

Wisconsin

University of Wisconsin Press
114 N. Murray St., Madison, WI 53715

THE WORLD WE CREATED AT HAMILTON HIGH
Gerald Grant

"A penetrating book that builds on the enduring insights of Goodlad, Sizer, and Boyer and offers special inspirations of its own. Grant takes a new tack in evoking an American high school, creating a kind of social biography that traces the movements influencing the character of Hamilton High ... Written with deep feeling for the noble possibilities in the profession, it should reach all but the most cynical." — *Kirkus Reviews*
$24.95

CHILDREN OF DIFFERENT WORLDS
The Formation of Social Behavior
Beatrice Blyth Whiting & Carolyn Pope Edwards

"A comprehensive theory that combine[s] developmental, cognitive, and social learning processes ... At last the biosocial bases of sex differences are beginning to look understandable."—Robert R. Sears, Stanford University $35.00

THE METRONOMIC SOCIETY
Natural Rhythms and Human Timetables
Michael Young

One of the most imaginative sociologists working today looks at the diverse ways in which society thinks about and measures time, and at the ways concepts of time influence society's patterns.
$25.00

FASTING GIRLS
The Emergence of Anorexia Nervosa as a Modern Disease
Joan Jacobs Brumberg

"A fascinating study of the role of eating disorders in the psychological lives of young women across the ages. The work is an outstanding example of the new scholarship that weds social history with developmental science to enhance our understanding of both individual and society." —Uri Bronfenbrenner, Cornell University $25.00

"Fading Away" (Courtesy, George Eastman House)

THEMATIC ORIGINS OF SCIENTIFIC THOUGHT
Kepler to Einstein
Revised Edition
Gerald Holton

"This book deserves to have a permanent influence on the work of historians, philosophers and sociologists of science, as well as on educationists and intellectual historians, but it can also be recommended to the general reader." —Stephen Toulmin, *Physics Today*
$25.00 cloth; $12.95 paper

The Chicago School

A Liberal Critique of Capitalism

Dennis Smith

Theoretical Traditions in the Social Sciences

This important new assessment of the Chicago School argues that both the strengths and weaknesses of Chicago sociology are derived from tensions within and between American capitalism and American liberalism. Smith offers a series of critical expositions of the school's central theorists — Albion Small, W.I. Thomas, Louis Wirth, Morris Janowitz, and others — and compares them with their contemporaries in Chicago (Veblen and Dewey) and Western Europe (The Frankfurt School).

**1988   288 pp.   ISBN 0-312-00384-6 $45.00 cl.
ISBN 0-312-00386-2 $14.95 pb.**

Women and the Life Cycle

Transitions and Turning Points

Edited by **Patricia Allatt, Teresa Keil, Alan Bryman** and **Bill Bytheway**

This selection of papers from the British Sociological Association Conference on the sociology of the life cycle focuses on the life course experience of women. The papers offer important insights into the question of gender in social relationships and current social policy debates about employment opportunities and welfare.

1988   220 pp.   ISBN 0-312-01343-4 $35.00

Television and Social Control

Mallory Wober and **Barrie Gunter**

This book asks whether society can and does control television. Can television's potentially harmful effects on society be nullified by regulating the range and scheduling of its contents? Will this be beneficial in the end? Wober and Gunter answer these questions and also examine the consequences of relinquishing all control over television.

1988   250 pp.   ISBN 0-312-01305-1 $29.95

Racism and Sexism

An Integrated Study

Edited by **Paula S. Rothenberg**

This anthology presents the first integrated study of racism and sexism within the concept of class. Combining a wide range of articles from sociology, history, psychology, philosophy, and literature, with illustrative material drawn from popular sources such as poems, essays, and legal decisions, the book relates the concepts of "racism", "sexism", and "class" to daily life. Clear and easy to read, the book makes a wide range of important material accessible to readers at all levels.

1987   464 pp.   ISBN 0-312-01217-9 $39.95

Class and Inequality

In Pre-Industrial, Capitalist and Communist Societies

M.B. Hamilton and **Maria Hirszowicz**

This book offers a comparative analysis of the most important debates relating to inequality and stratification. Covering all major concepts of the topic — including power, privilege, prestige, and social class — it provides an exposition and critique of the Marxian analysis of capitalism as well as discussions of such contemporary issues as the managerial revolution thesis and the expansion and diversification of the middle classes.

1987   320 pp.   ISBN 0-312-01222-5 $35.00

To order these or other St. Martin's Press Sociology titles, call toll-free 1-800-221-7945, or send a check or money order to: St. Martin's Press, Scholarly & Reference Division, 175 Fifth Avenue, New York, NY 10010

St. Martin's Press

Scholarly and Reference Division
175 Fifth Avenue
New York, New York 10010

Distributors of Manchester University Press, Berg Publishers, and Universe Books.

CURRENT SOCIOLOGY
La sociologie contemporaine

A Journal of the International Sociological
Association

Edited by **William Outhwaite** *University of Sussex*

Each issue of this unique journal is devoted to a comprehensive
Trend Report on a topic of interest to the international community of
sociologists. Authors review current trends and tendencies in all
areas of sociological work — theories, methods, concepts,
substantive research and national or regional developments. The aim
is to review new developments, to discuss controversies, and to
provide extensive bibliographies. From time to time, Commentaries
on Trend Reports are published in subsequent issues of the journal.

Recent reports have focused on the multinational version of social
science, different traditions of sociology and the human body, the
sociology of law, migration in Europe, the sociology of industrial and
post-industrial societies, and the sociology of humour and laughter.

Issues are published in French or English, but a text in one language
is always accompanied by an extensive resume in the other.
Current Sociology is an official journal of the International
Sociological Association. Its main aim is to review international
developments in the discipline and to provide a forum from which
professional sociologists from all countries can communicate with
the widest group of colleagues.

Current Sociology is published three times a year in Spring, Summer and
Winter
ISSN: 0011-3921

ASA Sponsored Publications...

Presidential Series

These annual volumes, published jointly with the American Sociological Association, are collections of original essays derives from presentations at ASA Annual Meetings. Titles include...

The Social Fabric   **$29.95 cloth**
James F. Short, Jr., Editor   **$14.95 paper**
    Composed of papers presented at the 1984 ASA Annual Meeting. The topic concerns research and theory regarding the nature of social order and social processes, and the contexts in which social life takes place. 360 pages. 1986. Order from: Sage Publications, 275 South Beverly Drive, Beverly Hills, CA 90212.

Gender and the Life Course   **$34.95 cloth**
Alice S. Rossi, Editor   **$15.95 paper**
    A state-of-the-art assessment of the best work currently being done on gender and age as a maturational factor. 1985. Order from: Aldine Publishing Company, 200 Saw Mill River Road, Hawthorne, NY 10532.

Sociological Theory and Research: A Critical Approach   **$34.95**
Hubert M. Blalock, Editor. 1980.
Societal Growth: Processes and Implications   **$19.95**
Amos H. Hawley, Editor. 1979.
    Order from: The Free Press, 866 Third Avenue, New York, NY 10022.

Individual members of the ASA are entitled to a special discount. Contact the ASA Executive Office for details.

Annual Review of Sociology

    Each year, the **Annual Review** presents a comprehensive overview of recent research on some of the most important areas of sociology. Prepared by recognized authorities, these Reviews allow the sociologist to stay abreast of developments in the field that may not fall within his/her area of specialization.

Volumes 1-11 (1975-85)—U.S. price, $27/volume; elsewhere, $30/volume
Volumes 12,13 (1986, 1987)—U.S. price, $31/volume; elsewhere, $34/volume
Volume 14 (1988)—U.S. price, $30; elsewhere, $43
(Add 6% of net price for volumes shipped to California)

    Order from Annual Reviews, Inc., 4139 El Camino Way, Palo Alto, CA 94306 Individual members of the ASA are entitled to a special discount. Contact the ASA Executive Office for details.

American Sociological Association
1722 N Street, N.W., Washington, D.C. 20036

Don't Miss This Valuable Resource!

Cumulative Index
of Sociology Journals

The Cumulative Index of Sociology Journals, published in early 1987, covers the 1971–1985 period. Compiled by Judith Lantz of Carbondale, it contains both an author and keyword or subject index. Books reviewed in any of the journals that are part of the Index are included in both the author and subject listings. The following ASA-sponsored publications are included in the 780-page Cumulative Index: American Sociological Review, The American Sociologist, Contemporary Sociology, Journal of Health and Social Behavior, Social Psychology Quarterly, Sociological Methodology, Sociological Theory, and Sociology of Education.

In addition, both the American Journal of Sociology and Social Forces have been incorporated into the index. Their inclusion represents the Association's first, and very preliminary, attempt to put together a single index of all major sociology journals. Future editions will include a larger number of journals.

The Cumulative Index is available to ASA members for $37.50, to non-members for $48.50 and to institutions for $65.00. Send prepaid orders to: ASA Executive Office, 1722 N Street, N.W., Washington, D.C. 20036. Act now to get this valuable new resource!

Please send me _____ copy (copies) of The Cumulative Index of Sociology Journals.

I have enclosed my check for _____ made out to ASA.

Name _____

Membership # _____

Address _____

_____

City _____ State _____ Zip _____

Please return this form and your check to the ASA Executive Office, 1722 N Street NW, Washington, DC 20036.

American Sociological Review

Volume 53 • June 1988 • Number 3

Official Journal of the American Sociological Association

AMERICAN SOCIOLOGICAL REVIEW

June 1988   Vol. 53, No. 3

Notice to Contributors

(Revised January 1987)

AUTHOR'S RESPONSIBILITY: *Manuscripts which do not adhere exactly to these suggestions will be immediately returned for revision. Only articles in the form specified below will be sent out for refereeing.*

The third edition of the *Publication Manual* of the American Psychological Association (available from A.P.A., P.O. Box 2710, Hyattsville, MD 20784) is very helpful in clarifying headings, abbreviations, table preparation, and the like.

Since manuscripts are evaluated as anonymous submissions, please keep identifying material out of the manuscript. Attach a cover page giving authorship, institutional affiliation, acknowledgments, and the date the article is submitted. Provide only the title as identification on the manuscript and abstract. Every effort should be made by the author to see that the manuscript itself contains no clues as to the author's identity.

Submission and Preparation of Manuscripts

1. Submit five (4) copies and retain the original for your files. Copies may be Xerox, mimeograph, or multilith, but not carbons. Manuscripts must be accompanied by a $10.00 manuscript submission fee, payable to the American Sociological Association.

2. Enclose a stamped self-addressed postcard for acknowledgment of manuscript receipt. Since manuscripts will not be returned by reviewers to the editorial office, the editorial office will not return them to authors.

3. All copy must be *typed, doublespaced* (including indented material, footnotes, and references) on 8½ by 11 inch white opaque paper. Line must not exceed six (6) inches. Margins must be a minimum of 1¼ inch on left and foot.

4. Include four (4) copies of an abstract of *no more* than 150 words.

5. Tables should be numbered consecutively throughout the article and typed on separate sheets at the end of the manuscript. Insert a location note at the appropriate place in the text, e.g., "Table 2 about here." Each table must include a descriptive title and headings to columns. Gather general footnotes to tables as *"Note:"* or *"Notes:,"* and use *a, b, c,* etc., for specific footnotes. Asterisks * and/or ** indicate significance at the 5 percent and 1 percent levels, respectively.

6. Illustrations submitted with the final draft must be of professional quality, ready for reproduction, executed on white paper or vellum, in black ink, with clear, medium weight, black lines and figures. All lettering on figures should be executed by an artist in pen and ink, by means of dry transfer letters, or by applying typeset material to the prepared artwork. Typewritten lettering should not appear in illustrations. Figures should be capable of legible reduction to a width size no larger than 5¾₆ inch (full page) or 2⁹⁄₁₆ inch (one column); they should be numbered consecutively, and the number and author's name should be penciled lightly on the back of each. All illustrations must have captions, which should not appear on the artwork but should be typed, double-spaced, on a sheet at the end of the manuscript.

7. Equations must be typed and important displayed equations identified by consecutive arabic numbers in parentheses on the right. Expressions should be aligned and compound subscripts and superscripts clearly marked if there is any potential for confusion. Indicate bold face characters by drawing a wavy line (a) under them; a single underline (a) means italic to a printer. Clarify all symbols with notes in the margins of the manuscript. Circle these and all other explanatory notes not intended for printing.

8. Three kinds of footnotes are possible each serving a different purpose:

   a. *Content footnotes.* Content footnotes are explanations or amplifications of the text. Because they are distracting to readers and *expensive* to include in printed material, an author should include important information in the text and omit irrelevant information. *Content footnotes will not be allowed generally.*

   Rather than footnoting long or complicated material, such as proofs or derivations unnecessary to the text, consider (i) indicating in a *short* footnote that the material is available from the author, (ii) depositing the material in a national retrieval center and including an appropriate footnote, or (iii) adding an appendix. If an appendix is used, the reference in text should read, for example: "(see Appendix A for complete derivation)."

   Text footnotes should be numbered consecutively throughout the article with superscript arabic

numerals. If after a footnote occurs it is later mentioned, use a parenthetical note "(see note 3)," rather than the superscript number.

b. *Reference footnotes*. Footnotes are used for reference purposes only to cite material of limited availability. Acceptable reference footnotes include (i) legal citations, which should follow the footnote style of *A Uniform System of Citation* (1967) published by the Harvard Law Review Association, (ii) copyright permission footnotes, (iii) unpublished works and works in progress.

c. *Table footnotes*. Table footnotes are appended *only* to a specific table. Footnotes to a table should be *lettered* consecutively *within* each table with superscript lowercase letters. (See 5.)

9. Acknowledgments, credits, and grant numbers are placed on the title page with an asterisk.

Reference Format

A. *In the text:* All source references are to be identified at the appropriate point in the text by the last name of the author, year of publication, and pagination where needed. Identify subsequent citations of the same source in the same way as the first. Examples follow:

1. If author's name is in the text, follow it with year in parentheses [". . . Duncan (1959) . . ."].

2. If author's name is not in the text, insert in parentheses the last name and year [". . . (Gouldner 1963) . . ."].

3. Pagination follows year of publication after a comma [". . . Kuhn (1970, p. 71)."].

4. Give both last names for dual authors. Give all last names on first citation in text for more than two authors; thereafter use "et al." in the text. When two authors have the same last names, include initials in the text. For institutional authorship, supply minimum identification from the beginning of the complete citation [". . . U.S. Bureau of the Census 1963, p. 117) . . ."].

5. Separate a series of references with semicolons and enclose them within a single pair of parentheses [". . . (Burgess 1968; Marwell et al. 1971, pp. 386–87; Cohen 1962) . . ."].

B. *In the appendix:* List all items alphabetically by author and, within author, by year of publication in an appendix titled "REFERENCES." The reference appendix must be complete and include all references in the text. The use of "et al." is not acceptable in the appendix; list the names of all authors using full first names. (See A.4. for text format.)

If there is more than one reference to the same author and year, distinguish them by the letters a, b, etc. added to the year [". . . Levy 1965a, p. 331) . . ."].

The first letter of each word in an article title should be capitalized. Titles of books and journals are printed in italics, so each word of the title should be underlined.

Give the publisher's name in as brief a form as is fully intelligible. For example, John A. Wiley and Sons should be "Wiley."

If the cited material is unpublished, use "forthcoming" with name of journal or publisher; otherwise use "unpublished."

Examples follow:

1. *Books:*   Mason, Karen O. 1974. *Women's Labor Force Participation and Fertility.* Research Triangle Park, NC: National Institutes of Health.

U.S. Bureau of the Census. 1960. *Characteristics of Population.* Vol. 1. Washington, D.C.: U.S. Government Printing Office.

2. *Periodicals:*   Conger, Rand D. Forthcoming. "The Effects of Positive Feedback on Direction and Amount of Verbalization in a Social Setting." *Pacific Sociological Review.*

Goodman, Leo A. 1974a. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." *Biometrika* 61:215–31.

_____. 1974b. "The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I—A Modified Latent Structure Approach." *American Journal of Sociology* 79:1179–1259.

3. *Collections:*   Clausen, John A. 1972. "The Life Course of Individuals." Pp. 457–514 in *Aging and Society,* vol. 3, *A Sociology of Age Stratification,* edited by M.W. Riley, M. Johnson, and A. Foner. New York: Russell Sage.

Elder, Glen H. 1975. "Age Differentiation and the Life Course." Pp. 165–90 in *Annual Review of Sociology,* vol. 1, edited by A. Inkeles, J. Coleman, and N. Smelser. Palo Alto, CA: Annual Reviews.

See 1986 and later issues for further examples.

ANNOUNCEMENT

Beginning July 1, 1988, manuscript submission fees for all ASA journals will be increased to $15.

CONTRIBUTORS TO THIS ISSUE

■ DAVID KNOKE (Incentives in Collective Action Organizations), Professor of Sociology at the University of Minnesota, has major interests in political and organizational sociology. With grants from the National Science Foundation, he and colleagues are studying the reliability of organizational measures and the network structure of the U.S. and West German labor policy domains. He has been awarded a Fulbright Senior Professor Award to spend the spring term, 1989, in Kiel working on the latter project. Edward O. Laumann and he recently published *The Organizational State: Social Choice in National Policy Domains* (University of Wisconsin Press).

■ MARK H. LAZERSON (Organizational Growth of Small Firms: An Outcome of Markets and Hierarchies?) is Jean Monnet Fellow for 1987–88 at the European University Institute, Florence, Italy. He is currently doing research on career patterns of women workers in clothing firms, and organizational strategies of small enterprises. Recent publications include an article in the *International Journal of the Sociology of Law* on dismissal legislation and factory conflict in Italian firms.

■ ROBERT C. LIEBMAN (Exploring the Social Sources of Denominationalism: Schisms in American Protestant Denominations, 1890–1980) is Assistant Professor of Sociology, Portland State University, where he is affiliated with the Systems Science Program. He is continuing his research on the familiarization of firms in nineteenth-century French weaving and extending his work on schisms to examine the process of schismogenesis. JOHN R. SUTTON is an Assistant Professor in the Sociology Department at the University of California, Santa Barbara. His recent research is concerned with the organizational dynamics of social control agencies and religious denominations. ROBERT WUTHNOW is Professor of Sociology at Princeton University. His recent books include *Meaning and Moral Order: Explorations in Cultural Analysis* (California, 1987) and *The Restructuring of American Religion: Society and Faith Since World War II* (Princeton, 1988). He is writing a book on ideology and social structure in the Reformation, the Enlightenment, and European Socialism.

■ ROBERT M. JIOBU (Ethnic Hegemony and the Japanese of California), Associate Professor, Department of Sociology, The Ohio State University, has just completed a monograph on ethnic groups in California. A forthcoming article in *Social Forces* concerns the career lengths of black, Hispanic, and white professional baseball players. He is currently working on a monograph about white and nonwhite groups, using data from the 1980 Census of Population.

■ WILLIAM G. ROY (Interlocking Directorates and Communities of Interest Among American Railroad Companies, 1905) is Associate Professor of Sociology at UCLA. He has published work on large-scale political and economic change in the late nineteenth and early twentieth century U.S. He is currently working on a study of the rise of American industrial corporations from 1880 to 1913, funded by the National Science Foundation. PHILIP BONACICH is Professor of Sociology at UCLA. He is currently doing experimental research on the ways in which communication networks can shape incentives to communicate or hoard information. This research is supported by the National Science Foundation. He is also working on a variety of other topics in the area of social networks.

■ DAVID L. TORRES (Professionalism, Variation, and Organizational Survival) is Assistant Professor, Management and Policy Department, University of Arizona. He is currently doing research on the population ecology of small businesses and small business strategies for survival. With J. Richard Harrison and Sal Masoud, he is also examining turnover and structure in top levels of management.

■ THOMAS SCHEFF (Shame and Conformity: The Deference-Emotion System) is Professor of Sociology at the University of California, Santa Barbara. His areas of interest include sociology of emotions, theory, mental illness, and in becoming less a specialist and more a generalist. He is the author of *Being Mentally Ill* (1984), Chair-elect of the Sociology of Emotions section of the ASA, and research consultant, Taskforce on Self-Esteem of the State of California.

■ JOEL A. DEVINE (Macroeconomic and Social-Control Policy Influences on Crime Rate Changes, 1948–1985) is Associate Professor, Department of Sociology, Tulane University. He is currently engaged in research on the distributive consequences of governmental policies and in writing a monograph on U.S. tax policies. JOSEPH F. SHELEY is Associate Professor and Chair, Department of Sociology, Tulane University. He is the author of America's "Crime Problem" (Wadsworth, 1985) and the forthcoming Handbook of Contemporary Criminology (Wadsworth). Currently, his research focuses on determinants of the behavior of law, crime and routine-activities patterns, and automobile theft. M. DWAYNE SMITH is Associate Professor of Sociology, Tulane University. His current research focuses on economic conditions and urban homicide rates. Also, in collaboration with Joel Devine and Joseph Sheley, he is exploring the effect of unemployment on crime rates among different segments of the U.S population.

■ KIRK R. WILLIAMS (The Social Production of Criminal Homicide: A Comparative Study of Disaggregated Rates in American Cities) is an Associate Professor in the Department of Sociology/Anthropology and a Research Associate in the Family Research Laboratory at the University of New Hampshire. He has published articles on inequality, violence, and social control. His current research activity includes writing a book on Family Homicide (with Murray Straus), collaborating on a study of physical and sexual aggression at the University of New Hampshire, and directing a national panel study of social control and violence in adult relationships, funded by the National Science Foundation. ROBERT L. FLEWELLING is currently a doctoral student and research assistant in the Department of Health Education, School of Public Health, at the University of North Carolina. His interests include the joint application of social science theory with epidemiology in developing and evaluating primary prevention programs and health promotion policy. Recent works deal with data quality and measurement issues in homicide research and the descriptive epidemiology of female-perpetrated homicide.

■ BRUCE LONDON (Urban Bias, Dependence, and Economic Stagnation in Noncore Nations) is Professor and Chair, Department of Sociology and Social Psychology, Florida Atlantic University. He has a long-standing interest in Third World urbanization and development. DAVID A. SMITH is Assistant Professor of Sociology at the University of California-Irvine. His research interests focus on comparative urbanization and development, and the structure of the world economy and its effect on national and local social change. Research on these subjects has recently appeared in Social Forces, Urban Affairs Quarterly, and International Studies Quarterly.

■ RONALD ANGEL (The Use of Subjective Information in Statistical Models) is Associate Professor of Sociology and Member, Institute for Health, Health Care Policy, and Aging Research, Rutgers University. His current research focuses on the health of Hispanics in the U.S. and on methodological issues in comparative research. WILLIAM GRONFEIN, Assistant Professor, Department of Sociology, Indiana University and Purdue University at Indianapolis, is currently conducting research on the determinants of the deinstitutionalization of the mentally ill and on the social-psychological and structural relationships linking paranoia, trust, and suspicion.

ERRATA

"Models for Comparing Mobility Tables: Toward Parsimony and Substance," by Kazuo Yamaguchi (*ASR* 52:482–94).

I would like to correct the following statements in my article. The second sentence of the abstract, which appears on page 482, describes novel aspects of modeling that are introduced in the paper and states, "These components include . . . (1) the strength of off-diagonal associations, (2) the strength of diagonal effects, and (3) the extent of structural mobility." Delete the second component, "(2) the strength of diagonal effects," and renumber (3) as (2).

On page 483, the last sentence of the first paragraph states, "But for models that characterize the pattern of association between rows and columns by multiple parameters, he [Hauser] allowed the parameters to vary freely with mobility tables." Replace this sentence with: "He employed two sets of models that respectively hypothesized variation and constancy across mobility tables in parameters that characterized off-diagonal association and compared the relative goodness of fit between the two. This comparison tests the homogeneity of off-diagonal association across mobility tables. But, except for those that employ only one parameter for off-diagonal association, the models with variation in parameters across mobility tables cannot characterize differences in the pattern of off-diagonal association either simply or very efficiently because they characterize the differences by several parameters. As for diagonal effects, however, Hauser employed models with efficient 1 d.f. contrasts in comparing mobility tables."

On page 485, second column, line 12, after ". . . by a single parameter," add: "The kind of models that permit a 1 d.f. contrast for the uniform difference in diagonal effects between two mobility tables were previously employed by Hauser (1984)."

Professor Robert M. Hauser pointed out to me that my paper understated the previous use of association models in comparative-mobility research and that two aspects of 1 d.f. contrasts that had been used in his and some others' work were inadequately represented in the paper. I regret any misleading statements or lack of discussion regarding the relevant literature of this issue.

One aspect pertains to the 1 d.f. comparison of diagonal effects. As I corrected above, Hauser (1984) previously introduced the same method of modeling diagonal effects as I do in my paper, which used a 1 d.f. contrast for comparing diagonal effects between mobility tables. In their analysis of a set of three-by-three mobility tables, Grusky and Hauser (1984) also used a model that assumed homogeneous row and column effects and a uniform (among statuses) diagonal effect and expressed the single parameter for the diagonal effect as a function of the exogenous variables that characterized nations. Hout (1984) used scaled association models to compare differences in the diagonal effect of each specified substantive dimension among groups and periods.

The other aspect pertains to the 1 d.f. comparison of off-diagonal association. Hout (1984) used scaled association models to assess differences in each specified substantive dimension of off-diagonal association among groups and periods. Grusky and Hauser (1984), with the model discussed above, expressed the latent distance between two status positions as a function of the exogenous variables. Although these two strategies for comparing off-diagonal association among mobility tables are different from the one I used, they are as efficient as mine because all strategies used 1 d.f. contrasts for comparing two mobility tables.

REFERENCES

Grusky, David B. and Robert M. Hauser, 1984. "Comparative Social Mobility Revisited: Models of Convergence and Divergence in the 16 Countries." *American Sociological Review* 49:19–38.

Hauser, Robert M. 1984. "Vertical Class Mobility in England, France, and Sweden." *Acta Sociologia* 27:87–110.

Hout, Michael. 1984. "Status, Autonomy, and Training in Occupational Mobility." *American Journal of Sociology* 89:1379–1409.

EDITOR'S COMMENT

ON REFERENCES

References are getting out of hand. Apparently, the fashion is spreading to have more and more of them. We have references for the history of an idea, for the "theory" we use, for mistakes that others have made, for the methods we use, for the omissions to treat a subject, and most of all, for our past publications. We even put references in the abstracts, in the footnotes, and in the tables and charts. Why this proliferation of references should mushroom is hard to fathom. Maybe it's because we have a fine historical sense not to omit anything that has occurred in our special subareas in the last fifty years. Or we need to demonstrate our profound knowledge of the literature. Perhaps, we want to make certain to include scholars who may turn out to be reviewers. Or possibly, our discipline is truly cumulative and we need to show how we are adding to the treasure trove. A recent author cited 195 references, but the mean hovers around fifty-five or sixty, or about one and one-half printed pages.

Clearly, adequate referencing is important. But it should be restrained and judicious. We don't need to reference the trinity (Marx, Durkheim, and Weber) to prove that the subject is important and part of the sociological heritage. Such referencing is decorative because we hardly ever cite the original date of the publication and the pages where the ideas appear. More important, reference-peppering in the first few sentences of a manuscript distracts the reader and interferes with the easy flow of ideas. I have seen as many as twenty references scattered within a single, tortured, compound sentence. I have to reread the sentence, skipping the references, to absorb the idea. It's not unusual for authors to devote one-third of their manuscripts pages to references. Such extremes are typically unnecessary and they increase our printing costs. Better to have more articles.

Can we work out a rational referencing policy? I don't know, but here are a few guidelines. If you must offer the uninitiated bibliographical help, refer to one or two sources that contain a good bibliography—such as a handbook, a source book, a chapter in the *Annual Review of Sociology,* an encyclopedia article, or even a chapter in a good text or monograph. If you have to demonstrate that important research has occurred over a lengthy period, cite an early, middle and late example (e.g., Spencer 1851; Durkheim 1898; Parsons 1941; and Zilch 1988). But these references are not worth much because they are unfocused. The references that count are the references that bear directly and specifically on the ideas you are investigating. A large number of references does not influence the savvy reviewers. The specific reference (including the page number) to the critically important point demonstrates your ability to converse with the informed elite circle. When reviewers complain that the author does not know the literature, they typically refer to a specific article that bears concretely on the point in question. Of course, the number of appropriate references varies with the subject under review. Most articles need fewer than thirty. A few need as many as fifty. The editor can't decide which reference should be omitted. Therefore, prudence calls for a little self-restraint on the part of the author.

MANUSCRIPTS FOR THE
ASA ROSE SOCIOLOGY SERIES

Manuscripts (100 to 300 typed pages) are solicited for publication in the *ASA Arnold and Caroline Rose Monograph Series.* The Series welcomes a variety of types of sociological work—qualitative or quantitative empirical studies, and theoretical or methodological treatises. An author should submit three copies of a manuscript for consideration to the Series Editor, Professor Teresa A. Sullivan, University of Texas, Population Research Center, 1800 Main, Austin, TX 78712.

INCENTIVES IN COLLECTIVE ACTION ORGANIZATIONS*

DAVID KNOKE

University of Minnesota

The effects of members' interests in incentives offered by collective-action organizations are examined with data from a national sample of American associations. Members expressed interests in six distinct dimensions underlying organizational-incentive systems, and these different aspects are specifically related to different types of member involvement, controlling for other personal and organizational attributes. Members with higher interests in normative and social inducements offered by their organizations are more likely to contribute time, money, and psychological commitment and to engage in internal participatory actions. Lobbying incentives are strongly related to external participation. Overtly utilitarian incentives such as material benefits, occupational rewards, and informational incentives are either unrelated to involvement or actually attract members unwilling to participate. The implications of these results for Olson's "by-product" or selective-incentive explanation of collective action are discussed.

The American Association of Retired Persons (AARP) is the largest nonprofit association in the United States. A recent advertisement in *Modern Maturity,* its four-color bimonthly magazine, offered in exchange for $5 annual dues not only the magazine, but also travel benefits, discount health insurance and pharmacy service, access to seven mutual funds, regular meetings at more than 3,000 local chapters of AARP, and a monthly news bulletin reporting "late-breaking stories and features on legislative issues that concern you." Potential supporters were urged to

"join AARP's more than 20 million members in defending the legislative interests of mature Americans everywhere." AARP offers a much wider variety of incentives than most American national associations. But the effectiveness of such inducements for engaging members' involvements has rarely been assessed systematically.

This paper undertakes such an investigation. First, it examines alternative theoretical explanations of individual motivations to contribute to collective-action organizations. Next, a review of relevant research findings from political science and sociological analyses of political decisions casts doubt on the efficacy of selective utilitarian goods to motivate individual contributions. An alternative approach is proposed in which differences in members' involvements are a function of multiple motivations and organizational incentives. Finally, tests of this approach use data from representative samples of members from 35 U.S. national associations.

THE PROBLEM OF COLLECTIVE ACTION

Collective action is a recurrent problem for citizens of mass societies. In a complex political economy, where huge organizational entities are the prime movers, individuals, by pooling their resources in organizations, can achieve objectives that they would be unable

to produce through their individual efforts. However, group members must trade personal control over their resources for the "multiplier effect" of collective action (Coleman 1973). For many, the benefits gained from collective action outweigh the loss in personal autonomy. This problem applies especially to groups that cannot compensate their participants for their involvement.

In contrast to firms and government agencies that rely on the market or taxing powers, collective actions by social movements and voluntary associations depend on resource inputs from a mass support base (Knoke and Prensky 1984). Occasionally, a wealthy "Sugar Daddy" will help to launch a new public interest group (Walker 1983); however, most collective-action organizations acquire the bulk of their resources from a critical mass of members or sympathetic "conscience constituents" (McCarthy and Zald 1977). Contributions take a variety of forms: money, in-kind services, time, psychological commitment. Once it has acquired control over such resources, an organization may allocate them to three basic types of collective actions or goals: (1) direct material services to its members; (2) normative legitimation through information and public-relations programs; and (3) political efforts to influence public-policy decisions (Knoke 1985, 1986). These activities can serve as incentives or inducements to members and potential members to continue providing the organization with resources to undertake collective actions. The relative importance of different types of incentives in different types of associations for member involvement in collective action has been debated for more than two decades.

Do Public Goods Motivate Involvement?

Mancur Olson's *The Logic of Collective Action* (1965) argued that public goods are inadequate for motivating individuals' contribution to collective actions. In Samuelson's classic definition, public goods "are enjoyed in common in the sense that each individual's consumption of such a good leads to no subtraction from any other individual's consumption of that good" (Samuelson 1954, p. 387). That is, "they must be available to everyone if they are available to anyone" (Olson 1965, p. 14). Many collective-action associations try to persuade the government to

produce public goods, such as tax rates, election reforms, import tariffs, professional status, or the right to own personal firearms, from whose enjoyment nonmembers cannot be excluded. Collective political actions, such as lobbying and publicity campaigns, are basic strategies of such organizations to secure public goods.

In contrast, private goods are consumed individually. Their methods of production allow a collectivity to restrict these benefits to persons possessing property rights, that is, membership standing (Riker and Ordeshook 1973, p. 245). Typical examples include an organization's newsletters and magazines, information services, group travel and insurance programs, workshops and social activities, and recognition awards. Collective-action organizations vary in the mixtures of incentives that they offer to their members, with some associations concentrating solely on direct services, and others (such as AARP) offering both services and public-policy lobbying efforts. The remarkable explosion in American advocacy groups in the past two decades has created hundreds of these mixed political economies at the national level (Schlozman and Tierney 1985; Berry 1977, 1984; Salisbury 1984).

For collective-action organizations that offer public goods as their sole incentives to potential contributors, both the "imperceptible effect" and "free rider" problems enter into an actor's calculations (Olson 1965). First, when a person's share of a collective good is small, especially in a large membership group, the amount of resources he or she should contribute will also be minuscule (Olson 1965, pp. 64–65). Second, because everyone shares in any public good that is produced by collective action, an actor should maximize her or his benefits by not contributing. But, if all members use the same calculus, all will withhold contributions. Therefore, "suboptimal" amounts of the collective good will be produced, i.e., none (Olson 1965, pp. 27–29). Olson's solution to these dilemmas was to posit the necessity for a group to offer selective utilitarian incentives (private goods) to supplement the public goods and thereby attract contributions for their production (1965, pp. 50–51). Under the so-called "by-product" theory of collective action, these private incentives become the main motivator of individual participation (Olson 1965, pp. 132–35). Subsequent to

Olson, various theoretical refinements of his collective-action model were proposed: imperfect information, the selective-inducement aspect of collective goods, and entrepreneurial emphasis on nonmaterial benefits (Moe 1980, pp. 18 and 142–44; Jenkins 1983); elaboration of the property space of utilitarian incentives (Zald and Jacobs 1978); the differential side effects of positive and negative incentives for collective action (Oliver 1980); the importance of leadership in fostering decision-maximizing behavior (Leibenstein 1978); and the primacy of collective bads and the costs to the individual of not contributing to collective actions (Mitchell 1979).

Many theorists were more concerned with qualifying the choice process than with conceptualizing alternative motives for participation. But adherence to strictly utilitarian cost-benefit calculations may unduly limit the scope of relevant collective-action organizations. At one point Olson argued, "Logically, the theory can cover all types of lobbies," including such noneconomic groups as those with "social, political, religious, or philanthropic objectives" (1965, p. 159). Then he exempted philanthropic and religious lobbies (p. 160), as well as lost-cause groups consisting of persons with "a low degree of rationality" (p. 161), and mass social movements whose adherents are "psychologically disturbed" (p. 162). Olson appeared to recognize that motivations in addition to utility-maximization may lead some persons to contribute to collective-action organizations. Rather than subsuming these motives under selective incentives through a tortured logic and thereby making his theory "no longer capable of empirical refutation" (Olson 1965, p. 160, fn. 91), he preferred to restrict severely the types of collectivities to which his by-product theory applies.

Alternative Bases of Motivation

Several schemes have suggested other bases of member motives and organizational incentives that cannot be reduced to utilitarian cost-benefit calculi: the instrumental-expressive dichotomy (Gordon and Babchuk 1959); Clark and Wilson's (1961) famous material-purposive-solidary incentive typology; Etzioni's (1975) utilitarian-normative-coercive bases of organizational compliance; the importance of solidarity and moral

principles (Fireman and Gamson 1979); and Knoke and Wright-Isak's (1982) rational choice, normative conformity, and affective bonding motivations. These typologies recognize that members vary enormously in their preferences for and responses to a diversity of incentives offered by collective-action organizations. Members' motives range from personal help with job searches, to social partying, to changing legal statutes that affect the lives of persons not belonging to the organization. Adherence to equity norms, standards of fairness, altruism, and emotional ties to persons and groups may all play some part in individual decisions to become involved in a collective-action association. Reducing such complexity to simple utilitarian calculi of net benefits minus costs distorts reality and ignores organizations' capacities to tailor their inducements to fit the diversity of members' interests. Efforts to test Olson's model and its alternatives have frequently shown that a variety of motives are important, as the following paragraphs report.

The empirical literature on organizational incentives is a steady accumulation of findings that refute the Olsonian emphasis on selective goods as essential for collective action. Wilson's (1973) case studies supported his contention that "political" associations, such as Americans for Democratic Action and the NAACP, rely on purposive as well as utilitarian incentives to attract and hold members. Several studies purported to test directly Olson's by-product thesis. Marsh (1976) found that firms join the Confederation of British Industry as much to obtain the public goods of government influence and because of their opposition to labor unions as to obtain its selective services. In Tillock and Morrison's (1979) survey of Zero Population Growth members, the virtual absence of private goods meant that most participants were attracted by the association's political- and public-education activities. Moe (1980), analyzing members of five large economic associations that engaged in legislative lobbying, found that their extensive packages of economic services

are widely used and valued by members, and they typically have greater inducement value than politics . . . [but] it is still true that a good many members join for political reasons—and, in fact, except in the Printers, most of the respondents in each group indicate that politics

plays a pivotal role in their decision to maintain membership (pp. 217–18).

A study of members' commitment to and participation in 32 Indianapolis social-influence associations found that purposive incentives were strongly related to involvements, while selective material inducements had almost no effect (Knoke and Wood 1981, pp. 50–69; Knoke 1981, 1983).

Godwin and Mitchell (1982) fitted six decision models to samples of persons living in an Oregon coastal commission's jurisdiction and to mail-survey data from members of five national environmental groups. They found that utilitarian models explained non-electoral participation, but needed to be extended to include norms of fairness and taking into account others' happiness. Marwell and Ames' (1979) field experiments with small groups of high school students also showed the importance of equity norms in overcoming free-rider tendencies:

> [N]ormative factors such as fairness seem to have strongly influenced economic decisions. It appears that subjects see the claims of groups for public goods as intrinsically normative in nature and their participation in the provision of goods as at least one major goal to be weighed in their matrix of motives (p. 1359).

A study of Three Mile Island area residents following the 1979 nuclear power plant accident found that self-interest was the main reason most people took a free ride on community activists' efforts (Walsh and Warland 1983). But activism was strongly related to pre-accident solidarity and ideology, net of discontent with the emergency. Oliver (1984) discovered that persons active in their Detroit neighborhood associations were more pessimistic than token members about their neighbors' willingness to make contributions. Thus, in small organizations, where individual contributions *do* make noticeable differences, the community pessimists allow others to take a free ride on their efforts. Hansen's (1985) times-series analysis of membership changes in three national interest groups indicated a strong response to the ebb and flow of collective political rather than selective benefits. Finally, research on participation in rebellious collective actions in New York and Hamburg, West Germany rejected private incentives in favor of "soft incentives" of public goods as the major

explanatory variables (Muller and Opp 1986; Opp 1986).

In conclusion, the evidence above suggests that collective-action organizations' members are motivated to become involved by their interests in a variety of incentives. The pure utility-maximizing egocentric actor of Olson's model does not fare well in the real world. Equity norms, interpersonal attachments, and political goals figure significantly in many participants' decisions to contribute. In particular, members' interests in public goods may be especially potent when associations try to attain their objectives by efforts to influence public authorities. Groups that lobby government officials for decisions (such as a farm commodity association or a women's rights organization) often mobilize their members' involvement in external influence activities. Because the goals sought are not personal but public, the primary inducement necessarily is an appeal to a public good rather than to selective benefits. In contrast, for groups that can supply their collective objectives primarily within their boundaries (e.g., a stamp collectors club), public goods seem far less relevant to engaging members' involvements. Hence, the political goals of collective-action organizations may prove a major factor in explaining members' responsiveness to incentives: public-goods incentives motivate members of associations having major political goals, while selective utilitarian inducements are more powerful for involving members of less political organizations.

The chief analytic task of this paper is to determine the relative impact of various types of incentives on members' involvements in their associations. How much importance do members actually assign to different public- and private-good incentives? How much effect do these different inducements have in explaining members' contributions, participation, and commitments? The next section proposes a framework for conceptualizing how organizational incentives and individual motives can be assessed.

A FRAMEWORK FOR ANALYSIS

This section elaborates a predisposition/opportunity perspective on the relationship between individual motives to contribute resources to organizations and organizational incentives to induce commitments from participants. The

origins of this framework were discussed by Knoke and Wright-Isak (1982), who drew on the micro-level behavioral model of Ajzen and Fishbein (1980) and the macro-level incentive typologies of Clark and Wilson (1961) and Etzioni (1975). The framework expects that a congruence or consistency between participant motivations and organizational inducements will elicit greater contributions of specific types of resources for collective action.

Member Motivations

Individual members of associations may be motivated (predisposed to act to attain valued goals) by three general processes (See Knoke and Wright-Isak [1982] for elaboration and citation of relevant sources):

1. Rational Choice: cost-benefit calculi that maximize the individual's expected utility.
2. Affective Bonding: emotional attachments to other persons and groups.
3. Normative Conformity: adherence to standards of conduct grounded in socially instilled values about principled behavior.

Neither affective-bonding nor normative-conformity motivations can be reduced to utilitarian cost-benefits, except by a tortuous logic that renders the latter concept universal and, therefore, useless for empirical work.

These three motivations are presumed jointly to affect basic decisions about an individual's involvement in collective action: whether to join an association, whether to remain a member, how much to participate in collective actions, and what amount of personal resources to contribute to the collectivity. The initial assumption is that each component independently affects these behavioral decisions, yielding the general multimotivational specification:

$$D = f(R) \cdot g(A) \cdot h(N)$$

that is, a person's involvement decision (D) is some function of each of three motivational predispositions to act—R (rational choice), A (affective bonding), and N (normative conformity). The greater a person's predisposition on each type of motive, the higher the probability of involvement in the organization. Beyond this general specification, we lack theoretical or empirical guidance about the functional forms of the relationship

between motives and decisions. Complex and nonlinear processes (e.g., exponential, logarithmic, threshold effects, feedback loops) may complicate the relationships. However, the purpose of this paper is not to estimate directly the impact of motives on behavior, but to use this general framework to understand how members respond to organizational incentives. To understand the connections, we must conceptualize organizational-incentive opportunities as mirroring the individual motivational predispositions.

Organizational-Incentive Systems

By offering members a variety of inducements, organizations try to attract sufficient resources to undertake collective actions. Three generic types of inducements comprise associations' incentive systems:

1. Utilitarian Incentives: private goods in the form of direct services to members that are consumed individually.
2. Social Incentives: jointly coordinated social and recreational activities whose enjoyment is also restricted to the membership.
3. Normative Incentives: primarily public goods that require collective efforts to influence governmental policy makers.

Associations vary markedly in the extent to which they combine these three types of incentives. Knoke and Adams (1987), analyzing part of the data used in this paper, found that 63 percent of all American national associations offered their members at least two of the three types of incentives, 30 percent specialized in only one type of incentive, while only 7 percent offered none. A multiple discriminant analysis revealed that the types of incentives provided were strongly related to organizations' formal goals, as well as to such structural features as size, division of labor, and environmental complexity. This paper seeks to extend the analysis by relating association-incentive systems to members' involvements. Specifically, the more complex an association's incentive system (i.e., the more diverse incentives it offers to its members), the more willing are members to relinquish personal resources to the organization. Variety being the spice of organizational life, a wide range of inducements attracts members with heterogeneous motivations or involvement.

The parallel between member motivations

and association incentive offerings is evident and intended: utilitarian services appeal to persons' rational-choice motives; social-recreational activities elicit involvement on the basis of affective bonding; and public policy-influence efforts (lobbying) attract persons with strong normative-conformity predispositions. A person's involvement in a collective-action organization depends on his or her responsiveness to the particular combination of inducements offered by the organization. A basic congruence relationship can be expected: the more closely a member's interests match an organization's inducements, the more she or he will contribute time, money, psychological commitment, participation, and other valued personal resources to the group.

Different types of incentives are expected to be more effective in eliciting different types of member involvements. The underlying principle is again a congruence between the types of inducements provided, member motivations, and their involvements. Persons whose primary interests are the utilitarian-selective goods and services would contribute mainly financial resources but would avoid participation or deep psychological attachments to the group. Members whose concerns center on affective social and recreational benefits are likely to respond with high rates of internal participation in those events or groups that involve interpersonal contacts, such as attending meetings and working on group projects. On the other hand, where public-goods benefits are a dominant motivation, the organizational provision of lobbying incentives can be expected to mobilize greater involvement in external affairs involving contacts with government officials about issues of concern to the association. These expectations about the impact of specific inducements on particular types of involvements are *ceteris paribus* relationships. That is, they are expected to persist even after controlling for organizational and personal factors that may affect contributions, participation, and psychological attachments to collective-action organizations. To test these hypotheses, quantitative data from a wide range of associations and their members were collected.

THE NATIONAL ASSOCIATION STUDY

This analysis uses data from the National Association Study (NAS) of the political economies of American collective-action organizations. Complete research design details appear in Knoke (1987) and the data in Knoke (1988c). A three-stage sampling procedure was followed. In the first stage, a stratified random sample of U.S. national mass-membership organizations was drawn from a master list.[1] Telephone interviews with leaders collected data on basic organizational structures and activities for 459 associations. In the second stage, 47 associations were sampled proportional to membership size from three strata: professional societies, recreational organizations, and women's associations.[2] The third stage involved mailing questionnaires to samples of each association's members. Letters were sent to the national organization leaders, inviting their participation in return for data sharing. Of the 24 professional societies, 20 permitted membership samples to be drawn, 2 refused, and 2 did not maintain a national centralized mailing list. Of the 15 recreational organiza-

---

[1] This list of 13,000 associations was compiled from three compendia (Colgate and Fowler 1983; Colgate and Evans 1981; and Akey 1983). Organizations were classified into five strata: trade associations, professional societies, labor unions, recreational associations, and a residual category. Samples of 60–120 organizations were selected at random from each stratum. In January and February 1984, hour-long telephone interviews were conducted with the executive director or chief elected official from each organization. Data were collected on a variety of organizational structures, financial activities, programs, and incentives. Interviews were completed with 459 informants, for an overall response rate of 92.7 percent.

[2] Because confidentiality was promised to the organizations in exchange for access to their membership lists, their names cannot be revealed here. However, some feel for the range of associations in each strata can be revealed by generic descriptions. The professional category includes organizations whose members belong to the same occupation: medical and health (4 associations), engineering (3), social scientists (3), biological and natural scientists (3), interior decorators, social workers, personnel administrators, child educators, computer analysts, musicians, and public relations officials. In the recreational category are clubs for sports car, airplane, bicycle, bird-watching, swimming, gardening, photography, and wilderness conservation enthusiasts. The six women's associations have members with special credentials or interests in issues involving women in colleges, labor unions, managerial, clerical, legal, and political positions.

tions, 9 participated in the project and 6 refused. Of women's associations, 6 of 8 participated (Knoke 1988a). Given the differential numbers and refusal rates among the three types of associations, somewhat larger membership samples were drawn from each of the recreational and women's organizations. A survey using self-administered mailback questionnaires began in September 1984 and was completed by May 1986. More than 14,000 questionnaires were sent to valid addresses, and 8,746 completed ones were returned: 3,879 from professional societies, 2,659 from recreational clubs, and 2,208 from women's association members. The overall response rate was 61.6 percent (professional 66.4 percent; recreational 59.7 percent; and women's 58 percent). Only 3 of the 35 organizations had member response rates below 50 percent.

RESULTS

The following subsections report the analyses of the NAS data sets (all analyses were conducted with an IBM PC-AT using SPSSPC+). Details of measurements and methods are explicated at relevant points. The substantive foci are the effects of incentive interests on members' involvement in activities of different types of organizations; the trade-off between services and lobbying; and the interaction of lobbying and communication on member participation in highly political and less-political organizations. The cumulative results paint a detailed portrait of the ways incentives affect members' involvements in collective-action organizations.

Member Interests in Incentives

Although a few prior theoretical speculations can be found, no one has attempted to measure the structure of members' incentive interests across a large number of collective-action organizations. Association members were presented a list of 23 incentives designed to tap several theoretical dimensions, and they were instructed:

Organizations offer their members various kinds of incentives. **How important** are each of the following organizational activities to you personally as a member of [organization name]? If an activity is not performed at present by the [org. name], just leave it unanswered and go on to the next item.

The four response categories ranged from "extremely important" to "not important" and were scored from 4 to 1, with 0 indicating that the item was not checked. Exploratory-factor analysis offers a reasonable method to reduce these data to more manageable proportions. Conceivably, a single factor could result on which all indicators load highly, implying a general dimension from high to low incentive evaluations. Alternatively, several substantively distinct dimensions might emerge, each one reflecting specialized evaluation of incentive importance. We can then create incentive scales by aggregating sets of indicators that tap underlying dimensions, to be used in subsequent analyses of members' involvements.

The incentive items were factor-analyzed using the principal components method and oblique rotation (no pair of factors correlated higher than ±.31). The six-factor results are reported in Table 1. The six factors that emerged exemplify the public/private good distinction. Four dimensions tap primarily selective goods or services available only to persons within the organization's boundaries: (1) Material Incentives consist of utilitarian commodities such as insurance plans, marketing and purchasing opportunities, and certification or licensing programs; (2) Information Incentives take the form of publications, data services, and research; (3) Occupational Incentives, particularly relevant for professional societies and unions, involve help with job searches, business-professional networking, and efforts to improve job conditions and incomes; (4) Social Incentives involve sociable or recreational activities, including friendship opportunities and cultural or artistic endeavors. The other two factors measure incentives that are more broadly focused on public or collective goods whose benefits may be shared by nonmembers: (5) Lobbying Incentives most notably include efforts to influence government policy decisions, as well as to affect nonmembers and the general public; (6) Normative Incentives combine "emphasizing the main principles and goals of the organization" with improving the prestige and status of the profession or the organization and also its image held by the public at large.[3]

---

[3] A subset of the incentive items had been included in the telephone interviews with the 459 organizational leaders. Using the same response format, they were asked, "How important do you believe the item is as an inducement to your

Table 1. Pattern.Loadings from Principal Components-Factor Analysis (Oblique Rotation) of Incentives for 8,746 Members of 35 Associations

Factors	I	II	III	IV	V	VI
Normative-Incentives Scale						
Enhancing the public status of profession/organization	.73	−.02	−.27	−.01	.03	.03
Educating general public about profession/organization	.62	.07	.08	.22	.25	.06
Stressing the general prestige of the organization	.65	.30	−.03	.06	−.10	−.03
Representing members in organization negotiations	.50	−.06	−.23	.09	−.03	−.29
Emphasizing main principles and goals of the organization	.41	.23	.17	.30	.26	.10
Social-Incentives Scale						
Offering social or recreational activities	.01	.83	.15	−.06	.02	−.14
Giving opportunities to form friendships	.12	.80	−.07	−.07	.06	.13
Enhancing members' cultural or artistic lives	−.03	.63	−.11	.03	−.04	−.14
Occupational-Incentives Scale						
Helping members with job searches	−.08	.04	−.82	.06	.12	−.01
Providing professional or business-contact opportunities	.20	−.00	−.77	−.12	.09	.06
Trying to improve members' incomes, economic conditions	.04	.01	−.62	.22	−.01	−.18
Lobbying-Incentives Scale						
Lobbying governments to influence legislation	.11	−.22	.06	.76	.14	−.08
Trying to change the lives or nonmembers	−.07	.18	−.16	.67	−.15	.15
Trying to change values and beliefs of the public	.22	−.04	.03	.65	.17	.07
Endorsing candidates for public office	−.02	−.03	−.09	.70	−.14	−.23
Conducting public relations or information programs	.40	.07	.13	.43	.23	.02
Information-Incentives Scale						
Publishing newsletters, magazines, journals	.09	−.08	−.03	−.22	.78	.02
Providing information or data services	−.07	.09	−.05	.05	.67	−.24
Sponsoring research activities	−.14	.07	−.18	.29	.55	−.03
Running seminars, conferences, and workshops	.22	.10	−.30	.01	.47	.19
Material-Incentives Scale						
Offering group insurance plans	−.01	.05	.06	.05	.11	−.79
Offering group purchasing or marketing opportunities	−.05	.27	−.11	.01	.05	−.66
Running certification or licensing program	.46	−.02	−.17	−.11	−.07	−.52
Eigenvalues	6.44	2.11	1.76	1.67	1.13	1.05
Percent of Variance	28.0%	9.2%	7.2%	7.2%	4.9%	4.6%

Six Incentive Scales were computed by summing responses to those items that load highly on a given factor, where 0 = not checked, 1 = not important, 2 = somewhat, 3 = very, and 4 = extremely important. Scale items were not weighted according to their factor loadings, as equal weighting makes no substantial difference in the measurement (Alwin 1973). Three other incentive measures were also constructed to control for the global structure of incentives. Incentive Diversity is the number of specific incentives (out of 23) for which a member expressed an extremely or very important interest (the mean of all members is 10.0). Incentive System Complexity is the number of dimensions (social, normative, or utilitarian) on which an organization offers high amounts of inducement, using information obtained in the

members?" An exploratory factor analysis of these data produced six factors with loadings remarkably similar to those obtained in the member data. Details appear in Knoke (1987) and are available from the author on request.

leader telephone interviews (see Knoke [1988b] for details). Incentive Incongruence is the mean squared difference across five basic types of incentives[4] between a member's interest level and the importance assigned to these inducements by the organization, as reported in the leader interview. A higher score indicates a larger discrepancy between member interests and organization offerings and hence more incongruence between predispositions and opportunities.

Effects on Member Involvements

A central question in this paper is how members' interests in organizational incentives affect their involvement in the collectivity. The general expectation is that persons with high interests in incentives are more likely to exchange their personal resources in

---

[4] Principles and goals; social-recreational; lobbying; general organizational principles; changing nonmembers' lives.

return for these benefits. However, as hypothesized above, involvement with a group takes many forms, and each form may be affected by different types of inducements. Specifically, interests in social-recreational incentives are most likely to stimulate internal participation, interests in normative inducements to mobilize external participation, and utilitarian-incentive interests to affect monetary contributions. Further, the impact of incentives on involvements may vary across types of associations, as functions of their collective purposes and objectives. Professional societies may attract members with interests in occupational benefits, recreational clubs those people seeking sociable contacts, and women's organizations feminists concerned about legislation. However, some types of inducements may have universal appeal regardless of specific association, for example, normative or informational incentives. In the absence of a compelling theory, the precise alignments of member-incentive interests and involvements within type of organization is primarily an empirical question. This section examines the effect of eight incentive measures on five member-involvement scales in three types of association, controlling for other organizational and personal characteristics.

Diverse measures of members' involvement in their organizations were collected. A factor analysis of 15 items found two distinct subsets: 10 activities that take place within the organization's boundaries and 5 activities that require external contacts.[5] Therefore, two scales—(1) Internal Participation and (2) External Participation—were constructed by averaging members' frequencies of performing these sets of activities (on a four-point scale from "never" to "regularly"). Three

single-item involvement indicators were also analyzed: (3) Give Time, the number of hours per month spent working for the group; (4) Give Money, the dollar amount, including dues, contributed annually to the organization; and (5) Commitment, measured by response to the question, "Personally, how committed do you feel to the [org. name]?", on a five-point scale from 'not at all' to 'very strongly committed'."

Tables 2, 3, and 4 report the regressions of the five involvement measures on the eight incentive measures,[6] while controlling for the effects of several variables.[7] Unstandardized

---

[5] The items in the Internal-Participation scale are phoned or contacted other members; ever held office in organization; voted in organizational elections; worked on special projects; tried to recruit new members, attended conferences or workshops; provided transportation to members; represented the organization to other groups; solicited donations for the organization; gave or loaned equipment and supplies. For the External-Participation scale, the items are contacted government officials; worked for a political candidate; signed a petition for the organization; wrote letters to papers or magazines; picketed or demonstrated. See Knoke (1987) for details of the index construction.

---

[6] Attempts to include the Incentive-Diversity measure foundered because of high collinearity with the six Incentive Scales in all three types of associations (multiple correlations exceeded .90). Unstable coefficients were produced whenever it was included. Incentive System Complexity had a similar multicollinearity problem among the recreational club members. Therefore, these predictor variables were omitted from the final equation specifications presented below.

[7] Briefly, these measures are:

1. Communication: the number of times per year the member reported communicating with organization officials and other members (contents unspecified);

2. Political Goals: a four-point scale from none to major goal of attempting to influence public-policy decisions of the government;

3. Social Change Goals: an average of affecting values of the larger society and attempting to change nonmembers' lives;

4. Sociable Goals: the importance of social or recreational objectives;

5. Bureaucracy: a three-level measure of national headquarters staff structure;

6. Decentralization: members' perception of the degree to which policymaking power is concentrated or dispersed, based on a six-item Likert scale;

7. Issue Interest Index: mean "level of member interest or concern" on a four-point scale for up to six recent issues faced by the organization;

8. Total Influence: the sum of member ratings of influence over policies and actions (from none to a very great deal) attributed to the president and to the general membership;

9. Influence Slope: the difference between presidential and general membership-influence ratings;

10. Years of Membership: number of years since joining the association;

11. Income: respondent's annual family income as midpoint of 12 categories;

Table 2. Regression of Organizational Involvement Measures on Incentive Scales for Professional Association Members (N=3,879)

Independent Variables	Internal Participation		External Participation		Give Money		Give Time		Commitment	
	B	b^*	B	b^*	B	b^*	B	b^*	B	b^*
Incentive Scales										
Normative	.12	.15*	.00	.00	.69	.00	5.65	.28*	.31	.32*
Lobbying	.00	.00	.11	.23*	−4.47	−.02	−2.52	−.11*	.06	.06*
Social	.11	.13*	−.02	−.04	12.71	.04	−.96	−.04	.09	.08*
Material	−.01	−.01	.03	.07*	4.24	.01	−1.52	−.07*	−.08	−.08*
Occupational	−.01	−.01	.00	.00	−4.47	−.02	.92	.05	−.05	−.05*
Informational	.06	.06*	−.01	−.02	−29.92	−.09*	−2.14	−.09*	.15	.13*
Complexity	.00	.00	−.00	−.01	22.39	.10*	.59	.04	.01	.02
Incongruence	.01	.03	−.00	−.01	−11.86	−.07*	.45	.04	.00	.00
Communication	.03	.43*	.01	.17*	3.19	.11*	.29	.15*	.01	.11*
Issue-interest index	.12	.13*	.04	.09*	24.22	.07*	−.40	−.02	.15	.14*
Political goals	−.03	−.05	.04	.13*	−32.42	−.15*	−.16	−.01	−.10	−.15*
Social-change goals	.06	.09*	−.01	−.05	15.10	.06	.29	.02	.07	.09*
Sociable goals	.11	.09*	.06	.08*	67.12	.14*	−1.05	−.03	.09	.08*
Bureaucracy	−.04	−.04	−.02	−.05	47.11	.14*	1.10	.05	.11	.10*
Decentralization	.03	.03	.01	.02	14.64	.04	.57	.02	.29	.23*
Total influence	.03	.05*	.01	.02	3.46	.02	−.25	−.02	.05	.09*
Influence slope	−.01	−.01	−.01	−.02	4.15	.02	−.16	−.01	−.01	−.01
Years of membership	.01	.11*	.00	.03	−.05	.00	−.07	−.04	.01	.11*
Income	.00	.05*	.00	.03	1.54	.15*	.01	.01	.00	.01
Education	.00	.01	.00	.02	−1.21	−.01	.22	.03	.01	.02
Gender	−.02	−.02	.00	.00	−12.98	−.03	−.69	−.02	−.08	−.05*
Age	.00	.00	.00	.07*	.53	.03	−.02	−.01	.00	.08*
Constant	.11	—	.58	—	−173.62	—	−5.04	—	.11	—
Multiple R^2_{adj}	.426*		.177*		.117*		.055*		.494*	

* $p<.001$.

coefficients permit direct comparisons of the magnitudes of variable effects across the same equations estimated for the three types of associations. Standardized coefficients show the relative importance of independent variables within the same equation. Given the large sample sizes, even very small estimated coefficients tend to be statistically significant. Only coefficients with probabilities less than .001 are indicated.

The patterns of significant and nonsignificant coefficients are generally consistent with expectations about incentive effects on member involvements. Consider first the three equations for internal participation. Across all three types of associations, members who express high interests in both normative and social incentives are much more likely to be active in their organizations' internal affairs. Members who are interested in lobbying and occupational benefits are either unaffected or

12. Education: the respondent's years of formal schooling;
13. Gender: 1 if the respondent is female, 0 if male;
14. Age: the respondent's age in years.

actually *less* likely to be active. The selective informational and material benefits produce only small effects. Material rewards slightly bolster recreational club members' participation (B = .07), but slightly diminish women's association involvements (B = −.09). The degree of incongruence between organizational offerings and member interests has no impact, while women's organizations that operate more complex incentive systems experience a small drop in members' internal activity (B = −.12).

The three equations for external participation are even simpler: for all types of associations, members with high interests in the public good of lobbying are more likely to be mobilized to contact government officials. This effect is about twice as strong for the women's associations (B = .18) as for the professional and recreational clubs (B = .11 and .09, respectively). Excepting a small positive effect of material inducements in professional societies, no other incentives are important for the recreational or professional societies. However, the women's association members are most complicated. Normative

Table 3. Regression of Organizational Involvement Measures on Incentive Scales for Recreational Association Members (N = 2,659)

Independent Variables	Internal Participation		External Participation		Give Money		Give Time		Commitment	
	B	b*	B	b*	B	b*	B	b*	B	b*
Incentive Scales										
Normative	.25	.26*	−.03	−.07	56.18	.10*	2.23	.08	.34	.36*
Lobbying	−.23	−.21*	.09	.20*	−83.44	−.13*	−3.52	−.12*	−.06	−.05
Social	.18	.19*	.01	.01	31.47	.06	1.44	.06	.05	.06*
Material	.07	.08*	.02	.04	21.82	.04	.71	.03	−.01	−.01
Occupational	−.20	−.22*	.01	.02	−41.59	−.08	.32	.01	−.16	−.17*
Informational	.00	.00	.02	.04	−1.55	.00	−2.21	−.08	.12	.12*
Complexity	▪		▪		▪		▪		▪	
Incongruence	−.02	−.03	−.01	−.03	−1.19	.00	−.30	−.02	−.04	−.07*
Communication	.04	.45*	.01	.27*	12.22	.23*	.55	.22*	.01	.13*
Issue-interest index	.12	.12*	.02	.04	30.54	.05	1.91	.07	.20	.20*
Political goals	.11	.19*	.02	.07	86.92	.25*	1.95	.12	.16	.27*
Social-change goals	−.04	−.04	.00	.00	−139.88	−.28*	−3.37	−.15	−.22	−.27
Sociable goals	.06	.10*	−.02	−.07	−75.93	−.23*	−.68	−.04	−.06	−.10*
Bureaucracy	−.25	−.29*	.02	.07	91.90	.18*	1.40	.06	.05	.06
Decentralization	−.16	−.13*	−.01	−.02	−41.31	−.06	.43	.01	.23	.20*
Total influence	.01	.02	−.01	−.02	−15.78	−.04	.05	.00	.05	.08*
Influence slope	−.03	−.06*	−.01	−.03	3.98	.01	.26	.02	.01	.02
Years of membership	.00	.05*	.00	.08*	.35	.01	.03	.01	.01	.12*
Income	.00	.01	.00	.02	.99	.05	−.02	−.03	−.00	−.01
Education	−.02	−.06*	−.00	−.02	−12.52	−.08*	.00	.00	−.01	−.05*
Gender	−.09	−.06*	−.02	−.04	−9.66	−.01	.35	.01	−.07	−.05*
Age	.00	−.05	.00	.03	−1.35	−.06	−.04	−.04	.00	.04
Constant	2.04	—	.84	—	481.00	—	2.99	—	1.58	—
Multiple R^2_{adj}	.560*		.154*		.172*		.083*		.519*	

▪ Omitted due to multicollinearity.
* $p < .001$.

inducements are as important as lobbying (both b*s are .18), while interest in occupational benefits discourages members' external involvement. In contrast to the internal-participation equation, a complex incentive system encourages women to become more active in external affairs (B = .10).

The equations for giving generalized time and money reveal heterogeneous patterns.[8] A

complex incentive system induces larger financial contributions from professionals (B = \$22.39) and women (B = \$21.31), while among recreational club members interests in normative incentives (B = \$56.18) elicit more money. The lobbying and personal-good incentives mostly have nonsignificant coefficients, casting doubt on a simple money-for-selective-goods exchange. Turning to giving time to the organization, public-good inducements appear to have contradictory effects: professional association members with high interests in normative benefits are more likely to spend more time on organization affairs, but those with strong lobbying interests are less likely to be involved. Lobbying also reduces involvement by recreational and women's association members. As the external-participation analysis revealed, people concerned about legislative issues are most likely to channel their efforts specifically toward contacting govern-

[8] Members were asked, "How much effect do you believe that your own dues and contributions (and only your own) have on the (org. name)'s success or failure in reaching its goals?" Less than 10 percent felt they had a "big effect," but these people gave an annual dues and contributions average of \$132, more than the \$84 from members who felt it had "some effect" (66 percent of respondents), and twice the \$63 given by those indicating "no effect" (25 percent of sample). Although these differences are statistically significant, so much within-category variation occurred that only one-third of 1 percent of the variance in money given was attributable to members' efficacy beliefs. The causal direction is ambiguous: do members give more if they feel efficacious, or does

the belief in a large impact follow from making a major donation?

Table 4. Regression of Organizational Involvement Measures on Incentive Scales for Women's Organization Members ($N = 2,208$)

Independent Variables	Internal Participation B	b*	External Participation B	b*	Give Money B	·b*	Give Time B	b*	Commitment B	b*
Incentive Scales										
Normative	.21	.18*	.17	.18*	21.39	.09	1.49	.09	.27	.31*
Lobbying	−.13	−.10*	.18	.18*	−1.38	−.01	−1.55	−.09*	.17	.17*
Social	.25	.25*	.03	.03	.94	.00	1.19	.08*	−.01	−.01
Material	−.09	−.07*	−.01	−.01	1.24	.00	−.76	−.04	−.07	−.07*
Occupational	−.11	−.12*	−.15	−.20*	−10.10	−.05	−.82	−.07	−.12	−.17*
Informational	.07	.06*	.03	.03	4.91	.02	1.55	.09*	.02	.02
Complexity	−.12	−.13*	.10	.12*	21.31	.10*	−.83	−.06	.06	.08*
Incongruence	−.02	−.03	.03	.05	2.49	.02	−.17	−.02	−.02	−.04
Communication	.03	.43*	.02	.25*	2.30	.15*	.37	.36*	.01	.16*
Issue-interest index	.14	.10*	.05	.04	11.18	.04	.40	.02	.17	.15*
Political goals	ª		ª		ª		ª		ª	
Social-change goals	.01	.00	−1.94	−.53*	−47.29	−.05	−6.44	−.11*	−.25	−.07
Sociable goals	.51	.28*	.89	.57*	20.78	.05	3.58	.14*	.18	.12*
Bureaucracy	ª		ª		ª		ª		ª	
Decentralization	.00	.00	.05	.04	−1.19	.00	.54	.03	.20	.20*
Total influence	.02	.03	−.01	−.02	1.65	.01	−.33	−.03	.06	.11*
Influence slope	−.00	−.01	−.01	−.02	−.09	.00	.46	.05	.01	.01
Years of membership	.01	.07*	.00	.04	.74	.04	−.05	−.05	.01	.10*
Income	−.00	−.01	−.00	−.01	.83	.10*	.00	.00	−.00	−.01
Education	.00	.00	.05	.15*	3.67	.05	.07	.01	.02	.06
Gender	.22	.04	.18	.04	9.21	.01	−.74	−.01	.00	.00
Age	−.00	−.07*	.00	.08*	.06	.00	.03	.03	.00	.05
Constant	.78	—	3.95	—	−94.22	—	14.89	—	1.55	—
Multiple R^2_{adj}	.555*		.378*		.045*		.207*		.512*	

ª Omitted because non-variance in this sample.

* $p < .001$.

ment officials on behalf of their associations. General time contributions are boosted by social and informational incentives for women. All other types of inducements either have no effect or actually reduce their members' time donations.

Finally, the three equations for psychological commitment to the association concur that normative incentives, a public-good inducement, are far greater than any other in eliciting high levels of support for a collective-action organization. In all three types of association, the standardized normative-incentive coefficients are substantially larger than those for any other variables. Their standardized values average one-third of a standard deviation, almost twice as large as the next most important incentive variable. The selective goods of occupational and material benefits clearly reduce members' commitments, but the social and informational inducements bolster support within professional and recreational associations. In contrast, the public good of legislative lobbying raises commitment by professional and women's association members, while not affecting recreational club members' commitment. The complexity and incongruence variables have only minor or nonsignificant impacts on commitment.

The tables report the effects of all the control variables. Space prevents a thorough discussion of all their contributions to member involvement (see Knoke 1987). Communication clearly has the most consistently powerful effect in all equations: the more often a member communicates with officers and other members "about matters of all kinds," the more likely that member is to participate and become committed to the organization. Another robust factor is interest in organization issues, with positive coefficients in all but one equation (half of them significant). The higher a member's interest in organizational affairs, the greater her or his involvement. Other variables, such as organization goals or structures, produced more inconsistent effects.

According to the multiple R^2s in Tables 2, 3, and 4, the entire set of predictor variables

jointly explains substantial proportions of the variance in internal and external participation and commitment, but fares poorly in accounting for time or money given. These latter two dependent variables are probably measured too grossly to discriminate among a variety of intentions people have when donating their effort and money (the latter mixes both mandatory dues and additional contributions). When participation is separated into internal and external domains, members clearly target their involvements according to individual interests in different types of incentive benefits. For example, a member who is attracted by the public policy-influence goals of a group is more likely to engage in group-organized political actions than is a member whose main concerns are making new friends. Even the effect of normative incentives fluctuates markedly across the full set of member involvements.

Involvement in Political Organizations

The preceding analyses demonstrated heterogeneous incentive effects on member involvements in different types of associations, defined as the three sampling strata used in the NAS. However, these organizations are differentiated along another important dimension that was obscured in the analysis—the nature of their collective-action targets. As discussed above, American national associations differ in their ability to achieve their objectives on their own or through efforts to influence public authorities. This section investigates how this political dimension interacts with members' incentive interests and other organizational features to affect mobilization for collective action.

During the informant interviews, association leaders were asked whether efforts at "influencing public policy decisions of the government" were a "major, moderate, minor, or not a goal" of their organizations. A division of the sample into 15 highly political organizations with major policy-influence goals (4,399 sample members) and 20 less political organizations with less emphasis on policy goals (4,347 sample members) covaries strongly with their stress on lobbying incentives.[9] The leaders of 13 of

the 15 highly political associations felt that "lobbying to influence government legislation and regulation" was a "somewhat important" to "extremely important" incentive for their members, compared to only 3 of the 20 leaders of the less political association who considered lobbying incentives that important. Further, 14 of the 15 highly political organizations "sometimes" or "frequently" made their positions on policy issues known to the federal government, while only 4 of the 20 less-political groups did so. These contrasts were also reflected among the members' reasons for joining. The less political members were only slightly more likely to cite direct services to members as a major reason for joining (38 percent versus 32 percent for the highly political group members). However, the difference in the two types of associations among members mentioning lobbying as a major reason for joining was substantial: only 6 percent for the less-political and 35 percent for the highly political organizations.[10] Thus, a comparison of the two groups' members interests in and responsiveness to incentives seems justified.

Following Moe (1980), the membership questionnaires posed two hypothetical questions about the trade-off between lobbying and services:

1. "Suppose the [org. name] were to stop offering services to its members and only engage in lobbying the government on public policies. What would be your likely reaction?"
2. "Now suppose that the [org. name] only offered services to its members and did not attempt to lobby the government on public policies. What would be your likely reaction?"

Two choices were offered: "I would probably drop out" and "I would probably remain a member." (Fewer than 5 percent of the respondents did not answer both items, and they were excluded from this analysis.) The joint distribution of these two items, shown in

---

[9] Six of the twenty professional societies, three of the nine recreational clubs, and all six of the

women's associations had major public-policy influence goals.

[10] Furthermore, among those who joined for lobbying, the percentage agreeing that they felt "a sense of responsibility or obligation" to remain in the organization was somewhat higher for the members of highly political organizations (79 percent) than for members of the less political ones (60 percent).

Table 5, permits a classification into four types of respondents, revealing the relative salience of services and lobbying in members' calculi of loyalty to the organization. The organizations' degree of politicization is strongly related to the members' responses. Services predominate among the less political associations (70 percent would stay if only services were offered, compared to only 34 percent for the highly political groups). Lobbying is substantially more important among the highly political members, 21 percent of whom would stay if only lobbying were available, compared to only 2 percent of the less political respondents. The highly political organizations also seem to attract persons with greater intrinsic loyalty, to judge by the one-third who claimed they would remain even if *neither* type of incentive were continued, while just 5 percent of the less-political members avowed such commitment.[11]

Given the greater centrality of lobbying inducements among the members of highly political associations, we might expect lobbying incentives to have a stronger effect on the organizational involvement of these persons, relative to members of less political groups. Stated as a hypothesis, lobbying incentives interact with organizational politicization to increase members' participation in external actions. Comparing the effect of lobbying incentives on external participation, the coefficient should be larger for members of highly political organizations than for members of

Table 5. Response to Hypothetical Choice Between Services and Lobbying by Members of Organizations with High and Low Policy-Influence Goals

Respondent Would:	Highly Political Organizations	Less-Political Organizations	All Orgs.
Stay if only services offered	34%	70%	52%
Stay if only lobbying offered	21%	2%	12%
Stay only if both offered	10%	23%	7%
Stay even if neither offered	35%	5%	29%
Total	100%	100%	100%
(N)	(4,190)	(4,153)	(8,343)

less political groups. This differential effect of lobbying incentives can be further elaborated as a second-order interaction involving frequency of communication with organizational leaders. Although the content of communicative interaction was not measured, frequency of contact probably covaries with persuasive messages concerning the public-policy issues facing the organization. For members who already have a high level of interest in the lobbying objectives of the organization, frequent communication with leaders and other members can serve as a reinforcement. The predisposition to act on behalf of the group, as indexed by high interest in the lobbing incentives, may combine with internal-communication frequency to mobilize members for additional efforts at external-influence actions. Thus, an interaction term involving lobbying incentives and communication should produce a significant increment to external participation, over and above the additive main effects of the independent variables. A similar interaction process was uncovered in a study of social influence-association members in a midwestern city (Knoke 1982). Because public policy-influence actions are more prevalent in the highly political organizations, the interaction effect is expected to be stronger among members of those groups than among members of the less political organizations (i.e., a second-order interaction occurs).

The interaction term was measured by a dichotomous · variable coded "1" if the respondent expressed higher than average interest in the lobbying incentives and. also communicated during the past year with

---

[11] If the frequencies underlying the percentages in Table 5 are treated as a 2 × 2 crosstabulation, odds-ratios reveal the extent to which members are willing to trade-off services and lobbying efforts. Among the highly political organizations, this odds-ratio is 2.13, while for the less-political associations it is 1.65. Thus, although members of both types of groups threaten to drop out over the loss of one type of incentive but promise to stay despite the loss of another kind, the effect is almost a third greater (the ratio of odds-ratio = 1.29) for the highly political members. A log-linear model testing whether the lobbying-services relationship is identical within both types of organizations produced a likelihood-ratio chi-square of 1398.6 for d.f. = 3, leading to clear rejection. This interaction effect implies that members of highly political collective-action organizations hold distinct attitudes about the importance of direct services and lobbying in their attachments to their associations.

elected officials and members of his or her organization and coded "0" if otherwise. The members of highly political groups were more than twice as likely as those in less political associations to exhibit this combination (22.9 percent to 9.1 percent, respectively). Similarly, they scored significantly higher on the scales of communication, interest in lobbying incentives, and external participation. Table 6 shows the results of the multiple-regression equations estimated separately for the members of the two types of organizations.

The hypothesized interaction is supported. For members of both types of organizations, the main effects of lobbying incentives and communication and their interactions are all highly significant. While the two pairs of main-effect coefficients have roughly comparable magnitudes (especially comparing the standardized b^*s), the interaction term is markedly stronger for members of the highly political organizations. Indeed, the unstandardized B for the highly political groups (.32) is

three times the B-effect for the less-political groups (.11). Comparing standardized main and interaction effects within equations, the lobbying-communication interaction ($b^* = .24$) equals the main effects (.18 and .22) for the highly political organizations. For the less-political associations, the main effects are somewhat smaller (lobbying = .14, communication = .20), and the interaction is weaker ($b^* = .13$). As a final measure of the interaction-term magnitude, its inclusion increases explained variance (adjusted multiple R^2) among the highly political members by 3.9 percent (from .322 to .361) but by only 1.3 percent for the less political members (from .150 to .163). Thus, from every angle, the combination of high interest in lobbying incentives and contact with organization leaders generates a substantial boost in the external-influence actions among the members of highly political organizations and a similar, but much weaker, increase among less political group members.

Table 6. Coefficients from Regression of External Involvement on Incentive Scales and Other Measures for Highly Political and Less Political Organizations

Independent Variables	Unstandardized Coefficients (B)		Standardized Coefficients (b^*)	
	Highly Political	Less Political Organizations	Highly Political	Less Political Organizations
Incentive Measures				
Normative	.07	.00	.09*	.00
Lobbying	.15	.06	.18*	.14*
Social	.00	−.02	.00	−.06
Material	−.05	.03	−.05	.07*
Occupational	−.05	.01	−.07*	.03
Informational	.03	−.01	.04	−.03
Diversity	.00	.00	−.02	.02
Complexity	.02	.00	.03	−.01
Incongruence	.03	.00	.07*	−.01
Communication	.013	.006	.22*	.20*
Interaction	.32	.11	.24*	.13*
Control Variables				
Social-change goals	−.13	.01	−.07*	.03
Sociable goals	.10	.02	.11*	.07
Bureaucracy	.14	−.02	.06*	−.06
Decentralization	.03	.00	.03	.01
Issue index	.05	.03	.05*	.08*
Influence total	.01	.00	.03	.01
Influence slope	.00	−.01	.00	−.04
Personal Attributes				
Years membership	.00	.00	.07*	.00
Family income	.00	.00	.01	.01
Education	.01	.00	.05*	.01
Gender	.12	.00	.10*	.01
Age	.00	.00	.00	.09*
Multiple R^2_{adj}	.361*	.163*		
(N)	(4,399)	(4,347)		

* $p < .001$.

Neither incentive diversity nor complexity significantly affect external involvement, but incongruence has a surprising positive coefficient in the highly political-organization equation. This effect implies that the more a member's incentive interests diverge from those stressed by the organization, the *greater* the member's involvement in external activities. If this relationship is not an artifact of multicollinearity (see note 6), it suggests that political mobilization occurs more often among members who are least conforming to the full profile of their association's internal-incentive economy.

DISCUSSION

This analysis leaves little doubt that organizational incentives are central to association members' decisions about their involvements in collective-action organizations. Persons with higher interests in particular inducements offered by their organizations are more likely to engage in internal and external participatory actions, to be more highly committed to the organization, and to contribute more time and money, even controlling for other personal and organizational attributes. The incentive-action connections are discriminating—not just any kinds of incentives will do. Members discern six distinct dimensions underlying incentive systems, and these different aspects are specifically related to different types of member involvement or avoidance. This selectivity suggests that members are attracted to, or seek out, those inducements that are most closely related to their central interests in an organization. Members who desire interpersonal social benefits engage in internal activities that allow them frequent interaction with other persons. Members attracted by collective efforts to influence public policy are more readily mobilized for contacting government officials and publicizing the campaign for social change. General normative principles, prestige, and status enhancements are especially potent instigators of general commitment and internal participation.

Not all the theoretical expectations were confirmed. Normative incentives do not mobilize external participation in professional and recreational organizations (although they did for the highly politicized women's associations). Such overtly selective-utilitarian incentives as material rewards and

information benefits are often unrelated to involvement or actually attract members unwilling to participate in most organizations. The expected importance of occupational inducements for professional-association members is not present. These negative results reveal a much more complex patterning of members' attraction to organizational incentives than previously adduced, calling for reconsideration of existing theoretical explanations.

In particular, Mancur Olson's narrowly construed "by-product" or selective-incentive theory of collective action is cast into doubt. His theoretical model, premised on microeconomic assumptions about individuals' utilitarian cost-benefit calculations, requires substantial private-good incentives to overcome the tendency of public goods to induce free riding. But the present analysis reveals many precisely contrary relationships: purely selective benefits (with some exceptions for social-recreational and informational incentives) tend to attract apathetic members, while the public goods of normative inducements are among the strongest factors motivating member involvements. The quintessential public good of legislative lobbying is crucial to mobilizing external participation by members of all types of associations. These patterns are consistent with evidence from several case studies and experiments suggesting that equity concerns, fairness principles, and altruistic norms regarding the well-being of others (including nonmembers) are exceptionally powerful social forces shaping the collective-action decisions of individuals. In this study, members of all types of associations seem willing to engage in "sacrificial" actions whose results may well benefit others more than themselves. Not surprisingly, a large majority of members who participated in lobbying efforts also acknowledged that "contributing one's share to the organization is the fair and equitable thing to do."

Research of the past two decades has brought us full circle from Olson's effort to destroy the conventional interest-group conception of collective-action organizations. A more sophisticated conceptualization has emerged about what attracts people to organizations and engages their energies in collective affairs. In real associations, members display highly heterogeneous motives that respond to a variety of organizational incentives with different kinds and amounts of

involvement. Olson's notion that members could be motivated only by selective-good incentives failed to recognize that significant numbers are motivated by public-good inducements. Ironically, those members who desire only to purchase services and who make only the minimum financial contribution permit a reverse Olsonian free ride. In the aggregate, these passive members provide their association with sufficient resources to create and sustain a more elaborate incentive system. These additional incentives, especially the public goods that appeal to activist members, generate from the latter the necessary participation for the association to pursue its collective-action goals. Thus, the association enjoys a free ride on the mass of small contributors who provide the wherewithal to mobilize the coordinated energies of its activist minority.

The dominant emphasis in incentive analysis has been a market orientation that views members' concerns about benefits as preconceived instrumental interests, rather than as a dynamic process of interest formation through political struggle. An alternative view considers how both individuals and groups "in general participate not merely to meet preexisting ends, but also to constitute themselves, or to reaffirm themselves, as persons and groups with particular and desired attributes" (Bowles and Gintis 1986, p. 138). Preferences for various types of incentives and collective-action objectives are not simply imported into an associational marketplace but are continually formed and reformed through mutually determining interaction among leaders and members. Hence, the importance of communication linkages for stimulating member involvements. An important research agenda is to uncover the micro-foundations of incentive demands and supplies through ongoing political processes within the organization. The frame-alignment approach proposed by Snow, Rochford, Worden, and Benford (1986), specifying necessary ideological preconditions for collective-action mobilization, offers a good point of departure. Although the social movement organization emphasis on identity formation and expression may be more clear-cut than in the conventional interest associations considered in this paper, in either situation, neglecting political forces that shape associations' incentive systems inevitably creates an unbalanced emphasis on the economic half of organizational political economies.

The discovery of a sizable interaction effect involving lobbying incentives, communication with leaders, and external influence activities calls for further investigation. Organizations cannot conduct continual mobilization campaigns, and most leader-member contacts probably convey routine message contents. Even using the crude indicators of the NAS, however, the critical importance of personal contacts for external partisan activity is evident. What remains to be explained are the conditions under which mobilizing inducements are brought into play and the ways that leaders attempt to persuade members to follow them as well as members' efforts to influence their leaders' decisions. Future analysis requires detailed observations of actual mobilization campaigns as they unfold over time, such as the National Rifle Association's frequent letter-writing drives against gun-control legislation or the rarer street demonstrations by social movement organizations.

The analysis of incentive systems must be integrated into a larger framework of organizational political economy. Incentives are both commodities and symbols exchanged for members' contributions of personal resources to a collectivity. Although this analysis controlled for a few basic indicators of the organizational polity (member issue interest, decentralization, influence distributions), it is not exhaustive. Much remains to be learned about how collective-action organizations reach their decisions on how to allocate the resources they acquire from their incentive systems. Formal and substantive organizational democracy lies at the political heart of this explanation, and the polity's connections to the incentive economy should be the critical focus of future reports.

REFERENCES

Ajzen, Icek and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Akey, Denise, ed. 1983. *Encyclopedia of Associations*, 18th ed. Detroit: Gale Research Company.

Alwin, Duane. 1973. "The Use of Factor Analysis in the Construction of Linear Composites in Social Research." *Sociological Methods and Research* 2:191–214.

Berry, Jeffrey M. 1977. *Lobbying for the People:*

The Political Behavior of Public Interest Groups.
Princeton, NJ: Princeton University Press.

————. 1984. *The Interest Group Society.* Boston:
Little, Brown.

Bowles, Samuel and Herbert Gintis. 1986. *Democracy and Capitalism: Property, Community, and the Contradictions of Modern Social Thought.*
New York: Basic Books.

Clark, Peter B. and James Q. Wilson. 1961.
"Incentive Systems: A Theory of Organizations." *Administrative Science Quarterly*
6:129–66.

Coleman, James S. 1973. "Loss of Power."
American Sociological Review 38:1–17.

Colgate, Craig, Jr. and Laurie A. Evans, eds. 1981
National Recreational, Sporting and Hobby Organizations of the United States. 3rd Annual Edition. Washington, DC: Columbia Books.

Colgate, Craig, Jr. and Roberta L. Fowler, eds.,
1983. *National Trade and Professional Associations of the United States, 18th Annual Edition.*
Washington, DC: Columbia Books.

Etzioni, Amatai. 1975. *A Comparative Analysis of Complex Organizations.* New York: Free Press.

Fireman, Bruce and William A. Gamson. 1979.
"Utilitarian Logic in the Resource Mobilization Perspective." Pp. 8–44 in *The Dynamics of Social Movements,* edited by M. N. Zald and J. D. McCarthy. Cambridge, MA: Winthrop.

Godwin, Richard K. and Robert Cameron Mitchell. 1982. "Rational Models, Collective Goods and Nonelectoral Political Behavior." *Western Political Quarterly* 35:161–81.

Gordon, Charles W. and Nicholas Babchuk. 1959.
"A Typology of Voluntary Associations."
American Sociological Review 24:22–29.

Hansen, John Mark. 1985. "The Political Economy of Group Membership." *American Political Science Review* 79:79–96.

Jenkins, J. Craig. 1983. "Resource Mobilization Theory and the Study of Social Movements."
Annual Review of Sociology 9:527–53.

Knoke, David. 1981. "Commitment and Detachment in Voluntary Associations." *American Sociological Review* 46:141–58.

————. 1982. "Political Mobilization by Voluntary Associations." *Journal of Political and Military Sociology* 10:171–82.

————. 1983. "Organization Sponsorship and Influence Reputation of Social Influence Associations." *Social Forces* 61:1065–87.

————. 1985. "The Political Economies of Associations." *Research in Political Sociology*
1:211–42.

————. 1986. "Associations and Interest Groups."
Annual Review of Sociology 12:1–20.

————. 1987. *Collective Action Organizations: The Political Economies of American Associations.* Minneapolis, MN: mimeo.

————. 1988a. "The Mobilization of Members in Women's Associations." In *Women in 20th*

Century Politics, edited by P. Gurin and L. A. Tilly. New York: Russell Sage (forthcoming).

————. 1988b. "Resource Acquisition and Allocation in U.S. National Associations." Forthcoming in *Organizing for Change: Social Movement Organizations across Cultures,* edited by B. Klandermans. Greenwich, CT: JAI Press.

————. 1988c. *The National Association Study*
[MRDF]. Minneapolis: University of Minnesota Center for Social Research [producer]. Department of Sociology [distributor].

Knoke, David and Richard E. Adams. 1987. "The Incentive Systems of Associations." *Research in the Sociology of Organizations* 5:285–309.

Knoke, David and David Prensky. 1984. "What Relevance Do Organization Theories Have for Voluntary Associations?" *Social Science Quarterly* 65:3–20.

Knoke, David and James R. Wood. 1981.
Organized for Action: Commitment in Voluntary Associations. New Brunswick, NJ: Rutgers University Press.

Knoke, David and Christine Wright-Isak. 1982.
"Individual Motives and Organizational Incentive Systems." *Research in the Sociology of Organizations* 1:209–54.

Leibenstein, Harvey. 1978. "On the Basic Proposition of X-Efficiency Theory." *American Economic Review* 68:328–34.

McCarthy, John D. and Mayer N. Zald. 1977.
"Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82:1212–41.

Marsh, David. 1976. "On Joining Interest Groups: An Empirical Consideration of the Works of Mancur Olson, Jr." *British Journal of Sociology*
6:257–72.

Marwell, Gerald and Ruth E. Ames. 1979.
"Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free Rider Problem." *American Journal of Sociology*
84:1335–60.

Mitchell, Robert Cameron. 1979. "National Environmental Lobbies and the Apparent Illogic of Collective Action." Pp. 187–221 in *Collective Decision-Making,* edited by C. Russell. Baltimore: Johns Hopkins University Press.

Moe, Terry M. 1980. *The Organization of Interests: Incentives and the Internal Dynamics of Political Interest Groups.* Chicago: University of Chicago Press.

Muller, Edward N. and Karl-Dieter Opp. 1986.
"Rational Choice and Rebellious Collective Action." *American Political Science Review*
80:471–87.

Oliver, Pamela. 1980. "Rewards and Punishments as Selective Incentives for Collective Action: Theoretical Investigations." *American Journal of Sociology* 85:1356–75.

————. 1984. "Active and Token Contributors to

Local Collective Action." *American Sociological Review* 49:601–10.

Olson, Mancur, Jr. 1965. *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Opp, Karl-Dieter. 1986. "Soft Incentives and Collective Action: Participation, in the Anti-Nuclear Movement." *British Journal of Political Science* 16:87–112.

Riker, William H. and Peter E. Ordeshook. 1973. *An Introduction to Positive Political Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Salisbury, Robert H. 1984. "Interest Representation: The Dominance of Institutions." *American Political Science Review* 78:64–76.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *Review of Economics and Statistics* 36:387–90.

Schlozman, Kay L. and John T. Tierney. *1985. Organized Interests and American Democracy.* New York: Harper and Row.

Snow, David A., E. Burke Rochford, Jr., Steven K. Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization, and Movement Participation." *American Sociological Review* 51:464–81.

Tillock, Harriet and Denton E. Morrison. 1979. "Group Size and Contribution to Collective Action: A Test of Mancur Olson's Theory on Zero Population Growth, Inc." *Research in Social Movements, Conflict, and Change* 2:131–58.

Walker, Jack L. 1983. "The Origins and Maintenance of Interest Groups in America." *American Political Science Review* 77:390–406.

Walsh, Edward J. and Rex H. Warland. 1983. "Social Movement Involvement in the Wake of a Nuclear Accident: Activists and Free Riders in the TMI Area." *American Sociological Review* 48:764–80.

Wilson, James Q. 1973. *Political Organizations.* New York: Basic Books.

Zald, Mayer N. and David Jacobs. 1978. "Compliance/Incentive Classifications of Organizations: Underlying Dimensions." *Administration and Society* 9:403–24.

ORGANIZATIONAL GROWTH OF SMALL FIRMS: AN OUTCOME OF MARKETS AND HIERARCHIES?*

MARK H. LAZERSON

European University Institute

The success of small manufacturing firms is premised on their ability to develop alternative organizational strategies. This research, centered in Modena, a province of 600,000 located in the region of Emilia Romagna in north central Italy, demonstrates that when small companies expand through vertical and horizontal integration, they usually create other small firms that they control. This strategy preserves the advantages that Italian small firms enjoy in terms of state support, labor-market flexibility, and organizational efficiencies. Though this organizational form replaces market relations with bureaucratic relations, it bears little resemblance to the markets-versus-hierarchies theory of Oliver Williamson. Small firms integrate vertically and horizontally to insulate themselves from competition, not from the opportunistic practices of buyers and sellers. Indeed, greater reliance by small firms on bureaucratic relations is frequently compensated by their increased dependence on market relations through intensified subcontracting.

INTRODUCTION

The prolonged economic stagnation and widespread unemployment of the recent decade have stimulated an unexpected reappraisal of the place of the small manufacturing sector in advanced industrial societies; until recently, this sector was considered an impediment to modernization and development (Berger 1980a; 1981). However, the predicted demise of the small firm has not come to pass. In the U.S., where the number of small firms has recently swelled (Granovetter 1984), there are growing doubts about the capacity of large economic organizations to satisfy the needs of an advanced society (Piore and Sabel 1984).

Nowhere among the advanced capitalist countries has the challenge of small firms been more vigorous than in Italy, the fifth largest capitalist industrial power. There, firms employing fewer than 50 employees engage 49 percent of the work force, and the average manufacturing firm has only 9.19[1] employees, compared to an average of 59.9 employees in American manufacturing firms.[2] While the number of industrial firms has risen 40 percent since 1971, the average number of employees per firm has dropped from 10 to 7.9.[3] This increased density of small firms in the last 10 years is closely correlated with both rising wealth and technological progress. Only in Italy's poorest and most underdeveloped regions have small firms been declining in number (Weiss 1984, p. 221).

Admittedly, any explanation for the boom among Italian small firms should consider the broader macroeconomic influences (Minsky

[1] Source: *Annuario di statistico Italiano*. 1986. "6th censimento generale dell'industria, del commercio, dei servizi e dell'artigianato, 26 october 1981." Pp. 347–77. Rome: Istat.

[2] County Business Patterns, 1968–1977: Ten Year History. U.S. Bureau of the Census, July 1981.

[3] Source: *Annuario di statistico Italiano*. 1986. "6th censimento generale dell'industria, del commercio, dei servizi e dell'artigianato, 26 october 1981." p. 347. Rome: Istat. These data of average firm size differ slightly from the earlier one cited because they include the construction industry.

1985) and the reality that a vibrant industrial sector depends on both small and large firms (Oakey 1984). Nevertheless, there is growing evidence that the success of small enterprises also rests on alternative organizational strategies. Research centered in Emilia Romagna, a mixed agricultural-industrial region of north central Italy with four million inhabitants, has detailed how small firms have often achieved better economies of scale than large ones through the creation of specialized industrial districts where an agglomeration of producers in one industry work in close physical proximity (Brusco 1982; Sabel 1982; Solinas 1982; Piore and Sabel 1984). In this so-called Emilian model, the narrow division of labor common to large firms is rejected in favor of an organizational structure based on employees who perform a variety of different tasks (Rieser and Franchi 1986; Agenzia Industrial Italiana 1984).

However, it would be a mistake to conclude that the economic vitality of these small firms rests solely on the technical superiority of their individual strategies within the marketplace. Rather, the evidence seems to indicate that small-firm expansion depends on a particular social and political order. The continued existence of the extended family, in modified form (Pitkin 1985), provides a foundation for economic relations based on cooperation and trust. The importance of such nonmarket relationships for economic success also explains the heavy reliance ex--worker owners place on turning to friends and former colleagues in their search for employees. These social factors are reinforced by national and local policies grounded in tax, company, and labor-relations legislation that assist small firms. But, in addition to state intervention, there is a rich network of private economic associations and political organizations that have constructed an environment in which markets prosper by promoting cooperative behavior and by providing small firms with the infrastructural needs that they could not afford alone. To capture the complexity of these phenomena, there is a need for additional research on how sociological variables contribute to the existence of markets.

The Research

The centrality of the social environment to any analysis of markets emerges from my research on the creation of satellite-firms, a strategy used by small Emilian firms to expand while maintaining the legal and organizational structure of a small company. In this article, I focus on 15 small mechanical-engineering companies in Modena, a prosperous province of 600,000 people within the region of Emilia-Romagna. The data are based on plant visits and open-ended interviews with employers and business association and trade union representatives. The 15 firms were drawn from a 1985 quantitative survey of 219 Modena firms (Rieser and Franchi 1986), representing the universe of mechanical engineering firms with eight or more employees associated with the Modena branch of the National Confederation of Artisans (CNA), a small-business association that sponsored the research. Three other companies in the survey reported using satellite-firms but refused to cooperate with my investigation. Another company expanded through a merger. In Modena in 1985, the CNA represented 53 percent of all mechanical-engineering artisanal firms.

Satellite-Firms

The 15 companies formed partnerships or invested capital in other small firms to expand horizontally or vertically. The additional firm may have been newly formed for this purpose or have already been in existence. Though I use the term satellite-firm, the additional firm need not necessarily be dominated by the first firm; indeed, it is possible that the satellite-firm may eventually outgrow the spawning firm. The firms also remain strictly independent legal entities, with separate accounting procedures, work forces, and workshops. Satellite-firms attempt to preserve the special characteristics of small firms through loosely coupled units yet exploit some of the advantages offered by more diversified economic groups. These organizational developments differ in many respects with the theories of Oliver Williamson, who has argued that firms expand because of dysfunctions in market relations and the superiority of bureaucratic forms of management (1975). Small-firm expansion in my Modena survey actually resulted in a growth of both markets and hierarchies within the same group of firms, something not predicted by Williamson's dichotomous theory. Furthermore, my Modena data indicated that the role of law and labor relations is a major variable in

shaping the small firm's decision to expand. Issues of hierarchies and markets barely address this central issue, nor do they help analyze how social and political institutions eliminate some of the more negative aspects of markets, avoiding the need for bureaucratic forms of control.

To better situate the discussion of satellite-firms, my paper begins with a brief exegesis of the phenomenon of decentralized production and the relative decline of the large factory. I argue that, while this development is not easily reconcilable with justification for large firms provided by Chandler (1977), it can be subsumed within Williamson's argument. After an abbreviated synopsis of Williamson's principal claims and the position of his adversaries, I proceed with a description of the particular organizational and legal structure in which small Italian firms operate. I then analyze the field research on satellite-firms and small-firm strategy and compare my empirical findings with the theory presented by Williamson.

THE DECENTRALIZATION OF PRODUCTION

Many of the technical justifications for large factories have been undermined. Though small firms may be less efficient than mass production facilities, such processes do not dominate manufacturing. Evidence suggests that even in industrialized economies manufacturing employment is concentrated in small batch and project production (Sayer 1986, pp. 57, 70). For example, in the United States and the United Kingdom, 75 percent of manufacturing production (by value) is made by means of batches (Littler 1985, p.19). Rarely has manufacturing resembled the widget-tightening contortions of Charlie Chaplin's *Modern Times*. Centralized manufacturing sites may have once been a function of the steam engine and water wheel that required production to be physically linked to the power source, as in the English textile factories studied by Marx. But electricity and small electric motors now permit production to be widely dispersed (Brusco 1975). Telecommunications, improved transportation facilities, and lightweight microelectronics have further decentralized production.

There remain, however, the social explanations for large factories that Marxists have attributed to the capitalists' need to overcome worker resistance (Marglin 1974). But the twentieth-century rise of trade-union power in large industries has undercut many of the advantages that employers gained from that form of organization. Today evidence indicates that small factories are less strike-prone than large ones (Prais 1982). The advantages of smaller, decentralized firms have not been overlooked by large industrial organizations seeking to enhance their control over the external environment, and the widespread dispersion of production by large corporations around the world has forced governments, suppliers, and especially labor unions to retrench (Scott and Storper 1985).

EXPLANATIONS FOR THE GROWTH OF LARGE FIRMS

There is nothing contradictory about the superior efficiency of small, decentralized production units and the tendency of large firms to absorb smaller firms, if one assumes that production is not transferred to larger units. But Chandler's (1977) explanation of the concentration of industry in the hands of a few large corporations is based largely on a centralization of production to improve coordination and throughput speed, even though the small-firm organizational model has proved it is not always necessary or important. On the other hand, Williamson's (1975) theory of firm expansion, which highlights transaction costs rather than economies of scale or technological aspects of production, should be as applicable to the growth dynamics of both small and large firms.

Williamson explains the combined presence of large, complex firms and small firms in the economy as a reflection of the relative efficiencies of hierarchies and markets in reducing transaction costs, which are involved in economic exchanges with buyers, sellers, suppliers, and distributors. In cases of nonrepetitive transactions, markets will usually prevail. Here, markets are efficient because the buyer is not dependent on the seller as in long-term contracts, where the latter is prone to squeeze the former, or, in Williamson's terms, behave opportunistically. Markets also are more efficient in cases of standardized and noncomplex products. The market for these products is quite large and price comparisons are easy to make because quality can be instantly compared. The

market will also function quite well if the barriers to entry are low and the production technology well known.

But Williamson says that, in long-term economic relationships based on the buyer's specific needs and the seller's special knowledge of them, the transaction costs are high because the size of the market is limited. The buyer has to invest expensive resources to protect against the seller's opportunism, whether it be in legal fees, quality controls, or administrative costs to assure contractual performance. He also pays inflated prices. Thus, in the case where the buyer is tied to a few sellers (reduced to small-numbers bargaining) and where the items are not fungible because of asset specificity, it is more efficient for the buyer to swallow the seller's company. It eliminates both Williamson's joint problem of bounded rationality, caused by the buyer's lack of adequate information about the seller's operations, and of opportunism, which derives from the seller's ability to conceal information from the buyer. In the newly expanded firm, bureaucratic rule, backed by hierarchical power, will limit opportunism, and organizational integration will restore transparency to economic transactions.

Williamson's analysis has been extensively criticized on a variety of grounds. Transactional economic analysis fails to recognize that within one organization opportunistic behavior may actually be more difficult to monitor and more costly than in separate organizations (Perrow 1981). There is also a strong counterargument that loose coupling among independent economic organizations avoids the rigidities and expensive overhead associated with tightly coupled multifunctional firms (Perrow 1986). Williamson has also been taken to task for his neoclassical economic assumption that competitive market forces and organizational hierarchies discourage human beings from acting egoistically in economic transactions. Much economic activity is underpinned by long-established and stable networks of personal and affective relationships (Granovetter 1985). Part of the strength of the Japanese models may also be in those relations of reciprocity and trust that are interwoven with market transaction (Dore 1983). Williamson himself has recently recognized the importance of cultural and institutional forces (1985). Finally, economic efficiency may be a motive for firm expan-

sion, but there are numerous historical cases of how it has suffered when business hierarchies have been formed to monopolize markets (Du Boff and Herman 1980).

THE EMILIAN MODEL

The Emilian model of small, highly specialized industrial firms does not seem to conform to Williamson's explanation of business behavior, even though many of these firms are enmeshed in types of economic transactions that would make them ripe for hierarchical integration. Small-numbers bargaining is rife. The CNA survey of 219 firms revealed that 48.5 percent of Modena firms that supply other firms have 10 or fewer regular customers (Rieser and Franchi 1986, p. 25). The very specialized, nonstandardized production that allows Emilian producers a niche in the market clearly is asset-specific and often involves only a few producers, two elements that should aggravate the twin-headed problem of bounded rationality and opportunism. While it is true that some of the technologies employed are fairly simple, many of them described by Sabel (1982) are quite sophisticated. Market-entry costs for a small firm are lower than for a large firm, but acquiring a skilled work force and the necessary capital still takes many years, slowing entry into the marketplace.

It also seems that the dense web of market relations, often marked by long-established reciprocal relations and interfamily connections, is one of the strengths of the Emilian model (Bagnasco 1985, p. 26). Clustered in municipally sponsored industrial parks or districts, small firms cooperate with one another to perform a multiplicity of tasks that in the past a large factory may have completed under one roof (Solinas 1982; Becattini 1979). This spatial division of production (Marshall 1970) allows small firms to enjoy economies of scale "at the level of one or a very few machines, not whole factories" (Sabel 1982, p. 226). Efficiency is achieved by rapid responses to market needs, based on a flexible organizational structure and substantial subcontracting to other small firms to allow specialization and maximum use of existing plant and equipment (Brusco and Sabel 1981; Rieser and Franchi 1986). These strategies increase rather than decrease interfirm dependency. The market exchanges also accelerate the

exchange of information and stimulate innovation (Messori 1986, p. 418). The effectiveness of these arrangements seems to have reduced the need for firms to expand to protect themselves from the caprices of the external environment.

The Social and Political Context

The organization of small firm in Modena is strongly marked by the province's particular political and social characteristics. Indeed, Modena was selected because of its advanced small-firm sector, not because of its representativeness. Its exceptionalism is apparent when compared to other areas, e.g., Naples, where exploitation of labor and tax violations are widespread in small firms (Capecchi and Pugliese 1978; Botta, Fonte, Improta, Pugliese, and Ruggiero 1976), and the Veneto region, also prosperous but under Christian Democratic control and relatively union-free (Bagnasco and Trigilia 1984). In Modena almost 20 percent of employees in small firms are union members,[4] a proportion between two and four times higher than the Italian average. This relative union strength derives from the dominance of the Communist Party, which, with a near electoral majority (compared to less than 30 percent elsewhere in Italy), has governed both the province and region alone or in coalition with other parties since the end of World War II.

Approximately sixty percent of small firms in Modena are members of the CNA, a national organization founded in the immediate post-war period that represents 300,000 artisanal firms. The CNA has close organizational ties with the Italian Communist Party and, to a lesser extent, with the Socialist Party, and its membership is disproportionately concentrated in Emilia Romagna and Tuscany, where the Communist Party has been historically dominant.

The CNA, which also performs accounting and bookkeeping services for many of its member firms, self-polices the local and national wage scales negotiated with trade unions. The CNA sees violation of these rules, along with fiscal evasion and nonpayment of employment taxes, as a form of unfair competition among its member firms.

The local Communist Party's interclassist social formation, including both workers and artisans, also allows it to act as a social mediator. In the province of Modena, strong zoning controls and industrial parks assure affordable workshops and prevent real-estate speculation in commercial property. A specialized agency of Emilia Romagna, financed by the region and fees from businesses, promotes the exchange of information and resources among small enterprises. It also assists in developing networks for the commercialization of products produced by small firms.

THE ARTISANAL FIRM

Artisanal status in Italy is a legal classification based primarily on the size of the company's work force, not the form of its productive activity[5]. Artisanal firms are permitted to hire anywhere from 8 to 40 employees; the lower limit applies to transportation firms and the upper to handicraft firms. Most small manufacturing firms are restricted to 22 employees, or 18 if they do not employ apprentices. The limit includes any family members, with exclusions allowed only for one partner and any handicapped persons. The CNA survey reports that between 1981 and 1985 there was a slight increase in the number of employees per firm from 10.9 to 12.19 (Rieser and Franchi 1986). Employment levels, however, may be exceeded by 20 percent for a period of up to three months per year. Hairdressers, auto mechanics, electricians, and other service providers are also considered artisans.

Of the firms in the larger survey, 80 of the 219 firms surveyed were not strictly artisanal enterprises. Nevertheless, these firms were all small—only 18 had more than 20 employees, of which 7 had more than 30. Among the firms with satellites, 10 were artisanal firms. The data for the 15 firms surveyed is contained in Table 1.

Because of the legislative, ideological, and historical underpinnings of the artisanal firm, conceived of as standing between capital and labor (Cavazzuti 1978; Weiss 1984), the law mandates that the artisan "must commit a substantial portion of his own labor to the

Table 1. Company Description

Firm	Year Est.	Work Force	Company Form	Principal Product
BGHI	1967	28	L.P.	Machinery for broom making
GMB	1977	13	Artisan	Hydraulic compressors
AUTL	1972	11	Artisan	Robot and control panels
OO	1968	12	Artisan	Automatic gates
FSS	1971	11	L.P.	Plactic products
DMM	1963	9	Artisan	Industrial furnishings
TMM	1970	12	Artisan	Automated equipment
BT	1961	62	Share Co.	Ceramic ovens
GEC	1978	12	L. Part.	Lamps
CIS	1979	15	L. Part.	Machinery for ceramic industry
CISI	1974	15	Artisan	Electronic systems
ORM	1954	13	Artisan	Rebuilt engines
FROS	1970	14	Artisan	Metal finishing
RENF	1971	13	Artisan	Metal hardening
CEC	1978	12	Artisan	Plastic dies

Note: L.P. = Limited partnership
Share Co. = Incorporated Firm

productive process."[6] Completely mechanized firms are denied artisanal status, while firms that rely on assembly-line production are restricted to 12 employees. Partnerships and cooperatives are permitted artisanal status, but only if a majority of the owners are directly engaged in production. Limited partnerships and shareholder companies are excluded from artisanal status.

Though artisan-entrepreneurs may be registered owners of only one firm, the law permits them to invest or form partnership interest in other artisanal firms. Through this method, small producers can integrate their resources with those of other artisans to expand their range of services and production without losing the flexibility, informality, and low overhead of small firms.

State Incentives

The legal definition of smallness is also important because it is associated with special dispensations from labor laws as well as other benefits. In Italy, firms with fewer than 16 employees may dismiss employees with relative impunity, unlike larger firms that are subject to the Workers' Charter, which mandates substantial fines and the reinstatement of employees dismissed without justification (Lazerson 1985).[7] Regardless of this

labor law, the national contract that regulates labor relations in artisanal firms also allows small firms more latitude than industrial firms in dismissing employees or economic reasons. Union organizing activities are also relatively unprotected in firms with fewer than 16 employees.

Artisanal status confers substantial loan and tax advantages and reduced administrative expenses. Artisanal firms are entitled to obtain state-subsidized loans for machinery and workshops of up to 120 million lire ($92,300) at five percentage points below the prevailing interest rates. In Emilia Romagna, regional loans are also available for up to 60 million lire at about three points below prevailing rates. Though these loans are inadequate for established artisan manufacturing firms that need to modernize, they provide seed capital for one or two skilled workers who wish to become self-employed. These subsidized state loans equalled 64 percent of artisanal investment requirements between 1953 to 1976 and accounted for 75 percent of new artisanal firms established between 1961 and 1971 (Weiss 1984, pp. 226–27). Other incentives are exemptions from some property taxes, exemption from maintaining inventory records, and protection against bankruptcy proceedings. The artisan's share of social-insurance contributions for workers is 47 percent of the employee's gross wages, four points lower than for industrial firms. Artisans' contributions to their own pensions are also heavily subsidized by the state. Nevertheless, the evidence seems to

---

[6] "La legge quadro per l'artigianato." Law of August 8, 1985, number 443, article 2.

[7] "Statuto dei Lavoratori, Law No. 300 of May 1970.

indicate that artisans receive less support from the state than large industries and state-run firms, and that more state subsidies have generally flowed to the depressed south than to the small-firm industries of north central Italy (Forte, Bevolo, Clerico, and Rosso 1978; Ranci 1983).

An entrepreneur's decision to expand through a satellite-firm is ultimately determined by market factors. But market criteria are continually structured by state regulations that pattern the forms under which capital is aggregated and reproduced: company and bankruptcy laws, taxation and fiscal policies, and labor laws.

In artisanal firms, all of the personal property of the owners is at risk in case of a law suit. Though the numerous advantages of artisanal status, along with liability insurance, usually outweigh this risk, manufacturing firms with substantial sales and commercial activities may set up a limited corporation for those businesses that expose the firm to increased risks and preserve the already existing artisan structure for the productive realm. Incorporating the sales division as a limited company is often also necessary because artisanal status does not cover commercial activities that are more than incidental to the firm's principal business.

Labor relations is central to understanding the growth strategies of small firms, but not for the reasons usually offered. Both in Italy and abroad, especially in the United Kingdom, it is argued that the small firm's principal attraction to business is as a haven from unions (Rainnie 1985, pp. 148–50; Bin 1983). In Italy, industrial-relations developments of the late 1960s and 1970s left small firms in a privileged position, where they were both exempted from the Workers' Charter of 1970 and relatively untouched by an unprecedented wave of trade-union militancy. In perspective, however, labor relations, particularly exemption from protective labor-law legislation, appear to be only one factor among many for the fluorescence of the Emilian model (Murray 1983). In Modena, wages are higher than the national average (Brusco 1982), and, since 1979, employees in artisanal workshops have been guaranteed the same wage increases as those in large industrial firms.[8] My interviews with union

representatives indicated that small firms, especially in the mechanical-engineering sector with its heavily male work forces, often pay higher wages than large companies. However, lower wages prevail in the clothing and textile sector of Modena Romagna that employs mostly female workers (Solinas 1982). Evasion of social insurance by small employers is also uncommon in the province of Modena, unlike in southern Italy.[9]

Of perhaps greater relevance is the priority that small employers place on not hiring more than 15 employees, above which level the provisions of the Workers' Charter take effect. Of the firms studied, four had already surpassed that level. Between 1981 and 1985, one firm lost its exempt status when its work force expanded while two other firms were exempted after a decrease in size. In both cases, the drop in employment seemed unrelated to the law. One of the employers expected that improved economic conditions would soon permit an expansion of the work force, while the other employer intended to hire an additional employee that would expand the work force beyond 15. On the other hand, several employers were very conscious of the 15-employee limit and refused to exceed it. The fact that less than 15 percent of the 219 firms surveyed had more than 15 employees offers circumstantial evidence that the Workers' Charter is an important consideration in the decision to expand. However, the effective threshold of the law is really higher, since employers with fewer than 50 to 100 employees have considerable flexibility in dismissing employees and often are only obliged to pay monetary damages rather than reinstate the employee (Treu 1975, 1976; Centro Nazionale di Prevenzione e Difesa Sociale 1984). Trade unions also are more flexible in negotiating dismissals with smaller firms regulated by the Workers' Charter. Equally important is the presence of a trade union in the firm that can constrain the employer's decision to dismiss (Lazerson 1985). Trade unions represented employees in 11 of the 18 firms. Even in firms unrepresented by union, there seemed to be some evidence that trade unions' institutional power in Modena af-

---

[8] *Contratto Collettive Nazionale di Lavoro,* of

June 14, 1984 for the mechanical engineering firms of the Province of Modena.

[9] Interview with inspector of the Italian Department of Health and Welfare (INPS) for the province of Modena.

forded employees limited protection. One employer of a nonunion firm told me that he renegotiated the dismissal of two employees after the union had intervened.

Though employers clearly preferred a free hand to hire and fire, work-force turnover was under five percent annually and dismissals infrequent in both union and nonunion firms. Of the 14 firms, two had dismissed employees for economic reasons and two others for personal reasons. Of course, in a number of the firms there were also periods of informal layoffs and reductions in work hours. But employment stability was a major concern of the employers, a reflection of the relative skill of the work force. Since most employees were trained on the job, employers loathed losing employees for economic reasons. These costs also made them wary about expanding their work force, preferring to meet increased demand by subcontracting to other firms. This policy discouraged employees from seeking work in larger firms because it increased their job security in case of economic difficulties. The employers' widespread view that employment should be seen as permanent was best expressed by one employer who said that a firm should not be a "seaport where people come and go."

THE IMPORTANCE OF DIRECT SUPERVISION

The general pattern that emerged from the research indicated that avoiding unions, paying low wages, or firing employees were not policies of small Modena employers. Rather, their key concerns were to assure the continuance of personal ties between themselves and their employees without the mediation of managers. In only two of the fourteen firms did employees supervise other employees. Employers feared that the close relation with their employees, which were described as "familial" and "collaborative," would be undermined in a larger firm. Even one relatively large nonartisanal firm in the survey favored decentralized manufacturing sites in close proximity to one another to preserve small work groups. The emphasis on close ties between employer and employees was even present in firms where all employees were union members and strikes had occurred. Strikes were usually called to renew regional or national contracts rather than to protest local conditions. Indeed one of the few strikes that was aimed at the local

employer resulted from the union's and CNA's opposition to a factory-level contract that exceeded the recommended wage increase for the province. In any case, strikes would almost always be followed by employees working overtime to make up lost production, a practice opposed by unions in large firms.

Unionization in these small firms seemed more an affirmation of solidarity with the working class outside the plant rather than a claim to institutionalize an adverse relationship with the employer inside the plant. Only four of the thirteen firms with unionized work forces negotiated contracts at the firm level in addition to the national and regional labor accords that apply to all employees, whether union members or not. Of these four, only one was negotiated with the participation of the union. Employers opposed firm-level contracts because of a desire to maintain a direct relationship with their employees and avoid rigid work rules. Flexible work schedules, frequent job rotation, overtime work (often in excess of that contractually permitted), a six-day work week, individual bonuses for workers, and pay scales sometimes above the union standard were aspects of small-firm labor relations that conflicted with national union policies. But since these practices often meant higher wages, union members in small firms rarely objected. The handful of employees who attempted in the 1970s to challenge these practices and apply the rigid contract provisions of large firms were rapidly eased out of the firms I surveyed.

This absence of sharp social conflict between employees and their artisan-employers reflects the social origins of the artisans and the desire of skilled workers to become artisan-employers themselves. The overwhelming majority of artisans surveyed (Rieser and Franchi 1986, p. 17) had been previously employed as skilled workers in larger firms, though some had come from artisanal enterprises. Indeed, union-shop stewards surveyed in large firms in the province of Modena defined artisan-owners as members of the working class (Franchi and Rieser 1983). A similar identity of interests with small-firm owners combined with a hostility toward large industrialists was also expressed in surveys of workers in the neighboring Communist region of Tuscany (Bagnasco 1985, p. 34). Evidence of this social affinity

between artisans and their employees was reported in the CNA survey, which revealed that 47.3 percent of new partners hired between 1981 and 1986 came from within the firm.

The Importance of Personal Ties

The informal arrangements between owners and employees was also solidified by small firms' preference for hiring people recommended by other firms and friends. Employers in the CNA survey hired 42.9 percent of their employees in this way and only 16.3 percent from the state employment office, from which larger firms are required to draw many of their employees. Perhaps because of this reliance on a closed network of trusted people, employers did not rely on the oppressive forms of simple control that have been described by Edwards (1979). The dominant impression I gained was that low trust attitudes toward labor were absent (Fox 1974). Only a minority of firms used time clocks, and some that did required everyone, including the partners, to use them. Employers were often permitted some flexibility in working hours. Nor were piece rates or any work-incentive pay systems used, reflecting both the essentially nonrepetitive nature of the work and the employer's preference for self-motivated employees. In any case, the relative absence of white-collar employees and well-defined lines of authority in small firms would have made direct-control strategies ineffective (Friedman 1986).

While satellite-firms preserved the intimate work relation between owner and employee, they often required the selection of new partners to manage them, the necessity of "having someone there" as two partners of a firm expressed it. According to the CNA survey, new partners were added principally to help manage and operate the firm. The partner's contribution to firm capital and his specialized skills were no more than secondary considerations. Hired managers were normally rejected because of the importance placed on ownership. However, the selections of new partners was always problematic because of the question of trust. The fear of not finding reliable and agreeable partners often was reflected by artisan-owners' selection of relatives or work colleagues as partners. In most cases, artisan-owners pre-

ferred not to expand rather than risk adding a partner to share the new work obligations.

MARKET REASON FOR EXPANSION

Although the problem of labor relations and company law were common to all types of small firms contemplating expansion, the precise reasons for establishing satellite-firms depended on the market position of the company. In general, artisans and small firms fell into two categories: those that marketed finished products and those that were subcontractors for other firms, usually performing only one phase of the production process. Of the 219 firms surveyed, 54.3 percent performed primarily subcontract work, while the remainder produced mostly their own products. However, of the companies that established satellite-firms, only 3 out of 18, or 16 percent, were primarily subcontractors. In general, small firms with their own product lines were larger, employed more white-collar employees because of marketing and design needs, and were more likely to use subcontractors themselves. Indeed, this widespread use of subcontractors demonstrates that even within the small-firm sector, a core and dependent periphery is visible.

For subcontracting firms, cost reduction was the principal objective, and if they established satellite-firms, it was to satisfy that goal rather than to enhance their control over the market. A good example was provided by three independent subcontracting firms that performed different finishing operations on steel roller bearings for a large constructor of heavy machinery. In the past, the roller bearings would make three round-trips in and out of the large factory and to and from the three smaller firms. But the severe recession of 1981–1982 that intensified competition forced the three firms to establish a jointly owned storage firm to reduce transportation and inventory costs. The three round-trips were reduced to one, and the placement of a storage facility near two of the small firms allowed much of the transport to be handled by cheap hand-operated dollies. This particular form of vertical integration also sharply increased interfirm dependency.

In some cases, reliance on satellite-firms was required because small firms lacked well-established relationships, either because business transactions took place in unfamiliar regions or involved third parties that supplied

financing. Thus, a manufacturer of broommaking machines was forced to establish a small firm in Spain to assist in the servicing and marketing of its products there. The inability to directly control its foreign distributor nearly ruined one lamp manufacturer after its French sales representative suddenly went bankrupt, leaving the company without an outlet in the important French market. A manufacturer's link-up with financial services is sometimes required because the buyers are unknown to the financing agent, even if the latter has a privileged relationship with the seller. Integration of financing services also facilitates sales, especially when the buyer is located abroad. This was exemplified by a very prosperous and well-established ceramic-oven manufacturer that established a separate financing firm to assist its customers with loan and leasing arrangements.

The expansion strategies of small firms with their own product lines were conditioned by the widespread division of labor that marks subcontract work. Thus, in none of the companies I visited were satellite-firms initiated to expand existing production. In the face of the potential costs of increased labor and capital inputs, firm owners usually preferred to subcontract any excess production rather than expand. Often these tasks could be outsourced to other small or large firms in the area that might even be more price-competitive. Indeed, a firm's decision to create new satellite-firms was often premised on the elimination or reduction of some of its existing production facilities to replace low-value items with higher-value ones.

Critical in an enterprise's calculations was that the satellite-firm structure lead to greater control over the three key phases of production: design, final assembly, and marketing. These functions not only helped insulate firms from competitors but also guaranteed them substantial autonomy in the market place, an autonomy that rested considerably on greater freedom to use subcontractors. The possibility that this increased outsourcing of production might impose higher transaction costs appeared to be a minor consideration unless it compromised the strategic core of the firm's market position, e.g., if outsourcing production risked revealing product and design secrets to competitors with whom a firm might share the same subcontractor. But such sensitive production phases could be kept in-house or entrusted to subcontractors with whom a long relationship of trust had been established. Other problems of subcontracting involved quality control, particularly when final customers demanded more exacting warranties, and the additional accounting and bookkeeping expenses. However, most firm owners believed that the higher transaction costs resulting from accounting expenses for satellite and subcontracting firms were preferable to the less desirable alternative of not knowing the cost of each phase of production, a problem that often arose with internal systems of control.

Satellite-firms were useful vehicles to launch new products and facilitate access to new markets and technologies. One example of how the satellite-firm structure was used to restrict opportunism in strategic areas, yet not compromise the small firm's flexible structure, was provided by a machine shop with 12 employees. In the past, it had specialized in ceramic boring tools but now produced automated box handling and robotic equipment because of a severe crisis in the ceramic-tile industry. Though staffed with a skilled work force, the firm was dependent on an engineering firm located in the same building for design concepts. To preserve this relationship, the small firm bought a five percent interest in the engineering firm that gave it first right to produce any new invention. But the opportunity for a complete merger was rejected by the machine shop because of concern that it posed an unacceptable risk should the economic environment worsen. Here, the fear of opportunism and dependence on a single firm led to some integration and reduced dependence on the market, while the need to maintain loose-coupling limited the move toward hierarchies. A similar concern for assuring a flow of new products led a manufacturer of automated equipment used in the ceramic and tile industry to transform a once independent subcontractor into a satellite-firm. This association with an electronics firm with its own design capacity and contacts outside the ceramics industry promised to enhance the primary firm's market autonomy.

Satellite-firms also allow enterprises to penetrate new markets without risking their existing activities. Two of the firms studied started out designing and producing plastic injection molds and dies, only to later enter the more lucrative market of plastic fabrication. Plastics production is highly valued

because it permits a direct relationship with consumers and increases a firm's options. However, a die-making firm that takes the step toward the actual production of plastics risks losing its customers who are also in the business of making plastics. This risk is reduced if the plastics production is contained within a satellite-firm that appears to be formally independent.

Finally, satellite-firms at the retail level not only provide secure market outlets for finished goods, but also offer the opportunity to illegally escape taxes. If a manufacturer is linked with a consumer retail outlet, it facilitates evasion of the country's tax laws. In transactions of goods between companies, an official receipt is normally demanded by the buyer to deduct the cost of the purchased item against gross income for tax purposes. However, a retailer who sells primarily to consumers can easily conceal most of his income because sales are often made without receipts, despite the passage of a new law requiring them. The real costs of products can also be easily concealed in transactions with a satellite firm. In this way, one manufacturer avoided declaring about 25 percent of its income.

CONCLUSIONS

It is possible that the organizational development discussed here is only a reflection of the youth of the firms and that, as they mature, their satellite-firms will be eliminated by full-scale mergers. Ten years ago, none of these firms had established satellites. Nevertheless, the firms' motivations for integration and combination only infrequently paralleled those suggested by Williamson. On the contrary, there is evidence that the more continuous and long term the exchange relations between economic agents, the less likely is the need to formalize them legally. Certainly from the standpoint of manufacturing flexibility, production costs, and insulation from economic difficulties, market relations seemed most attractive. The exception occurred when transactions with the external environment touched on what I have defined as the strategic core or the commanding heights of the market—design, marketing, and assembly functions. But here the rationale for vertical integration and greater intrafirm coordination was not to reduce transaction costs but to obtain a more secure

position in the market. Even more incongruous with the dichotomous vision of hierarchies and markets was the fact that, within a single firm, movements toward hierarchy and away from markets in one area were often counter-balanced by a greater dependence on markets in other areas.

The other critical issue that the market-and-hierarchies distinction does not adequately address is that of the legal and social environment within which firms operate. Before firms choose markets or hierarchies, they are affected by state policies that promote certain forms of business combinations more than others, often for reasons that are socially and politically, rather than economically, motivated (Berger 1980b). The strategies of Italian small firms would appear very different if tomorrow the entire legal structure promoting artisanal firms was withdrawn. This is also true for the special labor laws that make some organizational choices more attractive than others.

Analysis of social formation must also consider those particular national structures that encourage cooperation and solidarity, values that are not equally distributed and nurtured across national boundaries. In my research in Modena, these institutions included the local administration, the Communist Party, and the CNA. But in the Veneto region north of Emilia Romagna, they also included the Christian Democratic party and its vast capillary network (Bagnasco and Trigilia 1984). These political and social institutions' encouragement of collaborative arrangements and continuous intervention in the market has been necessary to prevent capitalism from destroying the environment in which it flourishes (Polanyi 1957), an event that would perhaps leave us with only hierarchies and no markets.

REFERENCES

Agenzia Industrial Italiana. 1984. *Osservatorio sulla subfornitura: Ill rilevazione—Settembre, 1984* Torino: Agenzia Industriale Italiana.
Bagnasco, Arnaldo. 1985. "La costruzione sociale del mercato: strategie di impresa e esperimenti di scala in Italia." *Stato e Mercato* 13:9–45.
Bagnasco, Arnaldo and Carlo Trigilia. 1984. *Societa e politica nelle aree di piccola impresa: Il caso di Bassano.* Venice: Arsenale.
Becattini, Giacomo. 1979. "Dal 'settore' industriale al 'distretto'industriale: alcune considerazione sull'unita di indagine dell'economia indu-

striale." *Rivista di economia e politica industriale*
1:7–22.

Berger, Suzanné. 1980a. "Discontinuity. in the
Politics of Industrial Society." Pp. 132–49 in
*Dualism and Discontinuity in Industrial Socie-
ties,* edited by Suzanne Berger and Michael J.
Piore. Cambridge: Cambridge University Press.

————. 1980b. "The Traditional Sector in France
and Italy." Pp. 88–131 in *Dualism and
Discontinuity in Industrial Societies,* edited by
Suzanne Berger and Michael J. Piore. Cam-
bridge: Cambridge University Press.

————. 1981. "The Uses of the Traditional Sector
in Italy: Why Declining Classes Survive." Pp.
71–89 in *The Petite Bourgeoisie: Comparative
Studies of the Uneasy Stratum,* edited by Frank
Bechhofer and Brian Elliott. London: Macmil-
lan.

Bin, Mario. 1983. *La piccola impresa industriale:
Problemi giuridici e analisi economica.* Bolo-
gna: Mulino.

Botta, Paolo, Maria Fonte, Lucia Improta, Enrico
Pugliese, and Francesca Ruggiero. 1976. "La
struttura del settore caizaturiero a Napoli."
Inchiesta 23:61–83.

Brusco, Sebastiano. 1975. "Organizzazione del
lavoro e decentramento produttivo nel settore
metalmeccanico." Pp. 8–67 in *Sindacato e
piccola impresa: Strategia del capitale e azione
sindacale nel decentramento produttivo,* edited
by the FLM of Bergamo. Bari: De Donato.

————. 1982. "The Emilian Model: Productive
Decentralization and Social Integration." *Cam-
bridge Journal of Economics* 6:167–84.

Brusco, Sebastiano and Charles Sabel. 1981.
"Artisan Production and Economic Growth."
Pp. 99–113 in *The Dynamics of Labour Market
Segmentation,* edited by Frank Wilkinson.
London: Academic Press.

Capecchi, Vittorio and Enrico Pugliese. 1978.
"Bologna Napoli: due citta a confronto."
Inchiesta 35–36:3–54.

Cavazzuti, Franco. 1978. "Le Piccole Imprese."
Pp. 549–628 in *Trattato di diritto commerciale e
di diritto pubblico dell'economia,* vol. 2, edited
by Francesco Galgano. Padova: Cedam, 1978.

Centro Nazionale di Prevenzione e Difesa Sociale.
1984. "L'Applicazione dello statuto dei lavora-
tori dal 1974 al 1981: Ricerca Campionaria in
Otto Province," *Quaderni della giustizia*
38:68–103.

Chandler, Alfred. 1977. *The Visible Hand: The
Managerial Revolution in American Business.*
Cambridge: Harvard University Press.

Dore, Ronald. 1983. "Goodwill and the Spirit of
Market Capitalism." *British Journal of Sociol-
ogy* 34:459–82.

Du Boff, Richard and Edward Herman. 1980.
"Alfred Chandler's New Business History: A
Review." *Politics and Society* 10:81–110.

Edwards, Richard. 1979. *Contested Terrain: The*
*Transformation of the Work Place in the
Twentieth Century.* New York: Basic Books.

Forte, Franco, Maripina Bevolo, Giuseppe Cle-
rico, and Loretta Rosso. 1978. *La redistribu-
zione assistenziale.* Milan: Etas Libri.

Fox, Alan. 1974. *Beyond Contract; Work Power
and Trust Relations* London: Faber.

Franchi, Maura and Vittorio Rieser. 1983. *Espe-
rienza e cultura dei delegati: un indagine nella
realta metalmeccanica modenese.* Reggio Emi-
lia: Bonhoeffer Edizioni.

Friedman, Andrew. 1986. "Developing the Mana-
gerial Strategies Approach to the Labour
Process." *Capital & Class* 30:97–124.

Granovetter, Mark. 1984. "Small Is Bountiful:
Labor Markets and Establishment Size." *Amer-
ican Sociological Review* 49:323–34.

————. 1985. "Economic Action and Social
Structure: The Problem of Embeddedness."
American Journal of Sociology 91:481–510.

Lazerson, Mark. 1985. *Labor Conflict Within the
Shadow of the Law: The Italian Workers'
Charter of 1970.* Ph.D. diss., University of
Wisconsin-Madison.

Littler, Craig. 1985. "Taylorism, Fordism and Job
Design." Pp. 10–29 in *Job Redesign: Critical
Perspectives on the Labour Process,* edited by
David Knights, Hugh Willmott, and David
Collinson. Aldershot: Gower.

Marglin, Stephen. 1974. "What Do Bosses Do?
The Origins and Functions of Hierarchy in
Capitalist Production." *Review of Radical Polit-
ical Economics* 6:60–112.

Marshall, Alfred. [1923] 1970. *Industry and
Trade.* New York: Kelley.

Messori, Marcello. 1986. "Sistemi di imprese e
sviluppo meridionale. Un confronto fra due area
industriali." *Stato e Mercato* 18:403–31.

Minsky, Hyman. 1985. "Review: The Second
Industrial Divide." *Challenge* 28 (3): 60–64.

Murray, Fergus. 1983. "The Decentralization of
Production — The Decline of the Mass-Collective
Worker?" *Capital & Class* 19:74–99.

Oakey, Ray. 1984. *High Technology Small Firms:
Regional Development in Britain and the United
States.* London: Francis Pinter.

Perrow, Charles. 1981. "Markets, Hierarchies and
Hegemony: A Critique of Chandler and William-
son." Pp. 371–404 in *Perspectives on Organi-
zation Design and Behavior,* edited by Andrew
Van de Ven and William Joyce. New York:
Wiley.

————. 1986. "Economic Theories of Organiza-
tion." *Theory and Society* 15:11–45.

Piore, Michael and Charles Sabel. 1984. *The
Second Industrial Divide: Possibilities for Pros-
perity.* New York: Basic Books.

Pitkin, Donald. 1985. *The House that Giacomo
Built: History of an Italian Family 1898–1978.*
New York: Cambridge University Press.

Polanyi, Karl. 1957. *The Great Transformation:*

The Political and Economic Origins of Our Time. Boston: Beacon Press.

Prais, Sigbert Jon. 1982. "Strike Frequencies and Plant Size: A Comment on Swedish and United Kingdom Experiences." *British Journal of Industrial Relations* 20:101–8.

Rainnie Al. 1985. "Small Firms, Big Problems: The Political Economy of Small Businesses." *Capital & Class* 25:140–68.

Ranci, Pippo. 1983. *I trasferimenti dello stato a imprese industriali negli anni settanta*. Bologna: Il Mulino.

Rieser, Vittorio and Maura Franchi. 1986. *Innovazione tecnologica e mutamento oganizzativo nell'impresa artigiano: Una ricerca sull'artigianato metalmeccanico di produzione nella provincia di Modena*. Unpublished manuscript, along with additional survey results on file with the Confederazione Nazionale dell'Artigianato of Modena.

Sabel, Charles. 1982. *Work and Politics: The Division of Labor in Industry*. Cambridge: Cambridge University Press.

Sayer, Andrew. 1986. "New Developments in Manufacturing: The Just-in-Time System." *Capital & Class* 30:43–72.

Scott, Allan and Michael Storper. 1985. *Production, Work and Territory: The Geographical Anatomy of Industrial Capitalism:* London: Allen and Unwin.

Solinas, Giovanni. 1982. "Labor Market Segmentation and Workers' Careers: The Case of the Italian Knitwear Industry." *Cambridge Journal of Economics* 6:331–52.

Treu, Tiziano. 1975, 1976. *Sindacati e magistratura nei conflitti di lavoro,* vols. I, II. Bologna: Il Mulino.

Weiss, Linda. 1984. "The Italian State and Small Business." *European Journal of Sociology* 25:214–41.

Williamson, Oliver. 1975. *Markets and Hierarchies*. New York: Free Press.

———. 1985. *The Economic Institution of Capitalism—Firms, Markets, Relational Contracting*. New York: Free Press.

EXPLORING THE SOCIAL SOURCES OF DENOMINATIONALISM: SCHISMS IN AMERICAN PROTESTANT DENOMINATIONS, 1890–1980*

ROBERT C. LIEBMAN
Portland State University

JOHN R. SUTTON
University of California, Santa Barbara

ROBERT WUTHNOW
Princeton University

Schisms are a major source of new religious denominations in America, but have received little attention in the sociological literature. This study is critical of the conventional assumption that schisms arise primarily from internal doctrinal disputes. Drawing on the resource mobilization literature, we offer an alternative argument that vulnerability to schism is related to the organizational characteristics of denominations. We apply dynamic quantitative techniques to longitudinal data on Protestant denominations in the U.S. to test hypotheses about denominational centralization, linkages to the wider environment, and demographic characteristics. Findings suggest that the larger the denomination, the greater the tendency to schism; the size effect is inhibited, however, when denominations are linked to interorganizational federations.

Issues of denominationalism and sectarianism have long fascinated sociologists of organizations, collective behavior, and religion, but they have paid little attention to schisms. While schisms have been recognized as a major source of new religious denominations, empirical research on the determinants of schisms remains sparse. In this paper we present initial findings from a comprehensive analysis of schisms in four major denominational families from 1890 to 1980.

As background to our results, we refer to theoretical and practical arguments concerning the importance of schisms. Sociologists have argued, for example, that the clearest way to distinguish sects and cults is by regarding the former but not the latter as the product of

schisms (Stark and Bainbridge 1979). Sects, in short, arise by breaking away from established religious organizations. Others have shown that schisms constitute one of the significant ways by which religious organizations adapt to changing political and economic environments (Wuthnow 1986). At a more practical level, religious leaders themselves have been much concerned about minimizing the possibilities for schism within their own organizations. Their worries have not been ill-founded, as shown by the schismatic groups that have broken from the main bodies of Presbyterian and Methodist traditions in recent years in the wake of large-scale mergers or by fears voiced periodically that even an organization such as the Southern Baptist Convention may be torn asunder by schismatic strife.[1]

Given the theoretical and practical significance of religious schisms, we anticipated finding a large body of published literature on this subject. However, a computerized bibliographic search, card catalogs at several major research libraries, several recently published bibliographic guides to the scientific study of religion, and our own exploration of the major journals, books, denominational histories, and unpublished papers netted only a

[1] For a discussion of the changing conflicts and cleavages in denominations, see Wuthnow (1988, pp. 71–99, 132–72).

few studies. These were all case studies of schisms (see especially Dornbusch and Irle 1959; Tubeville 1949; Vrga and Fahey 1970; Harrison and Maniha 1978; Steed 1986). Not a single study attempted to develop a systematic quantitative set of data on schisms and their antecedents. Lacking systematic empirical data, none of these studies contains a focused theoretical discussion about the possible correlates of schisms.[2]

Nevertheless, it would be incorrect to suggest that the literature contains no clues about the possible origins of schisms. Conventional impressions tend to point toward both disputes over doctrine and problems of church governance and organization as the driving forces behind schisms. For example, a widespread view—popularized by reports on the New Christian Right in recent years—is that theologically conservative denominations are more prone to schisms than liberal ones. Conservatism, according to this view, generates dogmatism, which decreases tolerance of internal diversity, increasing the chances of dissident factions breaking away to form their own organizations.

Another impression—one that comes from the literature on churches and sects—is that schisms may originate as a kind of reaction against the seemingly inevitable tendency of sectlike groups to take on the characteristics of churchlike organizations. That is, there may be something about the sheer growth of religious organizations, or their tendency to become more bureaucratized and formalized, that leads minorities within these organizations to break away in pursuit of the original, purer values of their founders.

Still another impression—one we gained from talking with denominational bureaucrats during the early stages of our research—is that the potential for schisms is different across denominational families. Some denominational traditions, we were told, simply provide stronger precedents for breaking away. Or, as other leaders described it, some denominational families seem to be more rigid organizationally. Consequently, they experience the difficulties often encountered by movements in authoritarian states: for protest to be heard, some faction has to organize a palace revolution, and if the

revolution fails, the faction may be forced into exile or may voluntarily withdraw to form a separate organization. By implication, other denominational polities may have structures better suited to accommodating diversity and absorbing dissent.

The most influential study on this topic is H. Richard Niebuhr's *Social Sources of Denominationalism* (1929). Influenced by the work of Weber and Troeltsch, Niebuhr attacked the conventional interpretation that denominationalism is largely the product of doctrinal disputes, and argued instead that the fracturing of religious communities is driven by social factors—especially, in the American case, social differentiation in terms of race, class, and sectionalism. His work, especially his critique of doctrinal explanations, is a starting point for a sociological analysis of denominationalism, but for our purposes it is incomplete. Niebuhr's purpose was not to offer a theory of schisms, but rather to account for denominational diversity. His account focuses more on the societal factors associated with denominational diversity than on variation across denominations in vulnerability to schism and secession.

In this study, we follow Niebuhr in emphasizing the social, rather than purely doctrinal, sources of schisms. But, to develop hypotheses about why some denominations are more prone to schisms than others, we treat denominations as organizational systems, similar in many ways to firms, social movements, professional associations, and armies. We define a schism as the successful formation of a new denomination as a result of a break from a pre-existing denomination. Schisms, it seems to us, were probably affected by characteristics of the parent body in much the same way that challenging movements are affected by the larger organizational structures in which they appear. Thus, we draw on the resource-mobilization literature, particularly the work of Gamson (1975) and Zald and Berger (1978), to develop indicators of denominational structure that suggest the kind of resources and opportunities available to potential secessionists. We focus on the effects of centralization, size, and membership in interdenominational federation on vulnerability to schism. Our goal is to define and test a set of theoretically inspired hypotheses, but because this is an exploratory study, our indicators and results should be judged on their own merits, rather

---

[2] A brief introduction to the topic of schisms is presented in Wilson (1971) and Stark and Bainbridge (1985).

than being forced into a deductive theoretical model.

The value of our data is twofold. First, they cover all schisms occurring in the four largest Protestant denominational families in the United States—Baptists, Lutherans, Methodists, and Presbyterians—over the past century. Second, we are able to relate the incidence of schisms to a number of other organizational characteristics of the religious bodies in which these schisms appeared. Our results are sometimes predictable, sometimes counterintuitive. They are always subject to interpretation. This is partly due to the inevitable difficulties associated with making inferences from single indicators, and also to the fact that one can view the results from several different theoretical frameworks. We first present hypotheses and describe the means by which we gathered and coded our data. We then present the results of our analysis, using descriptive graphics and formal multivariate models. Finally, we speculate about how the results can be interpreted.

HYPOTHESES AND DATA

Hypotheses

Our null hypothesis is that the propensity to schism is determined by ideal factors such as doctrinal disputes or familial propensities. We lack reliable, fine-grained indicators of doctrinal differences or of the kinds of doctrinal controversies (if any) that were salient in specific schisms. As a rough proxy for doctrinal variation, we have grouped the denominations in our sample into families. Our strategy is, first, to estimate the degree to which schisms can be attributed to family membership; and second, to falsify this association by introducing indicators of the organizational structure of denominations. If a significant family effect persists when other factors are controlled, this will constitute an important challenge to our organizational argument.

The first set of denomination-specific factors concerns church governance: the centralization and formalization of denominational authority. First, we use the classic tripartite scheme of polity types, distinguishing among episcopal, presbyterian, and congregational polities. In episcopal structures, decision-making powers reside in a central

body of bishops; in presbyterian structures, authority resides in a body of local elders elected by lay members of the church; and in congregational structures, decision-making power is held by the congregations of individual churches. This conception of polity provides a good indicator of the centralization of formal authority and, hence, of control over resources. It reflects where both doctrinal authority and ownership of church property and financial assets are vested.[3] The more centralized the authority structure, the more difficult it should be for insurgent groups to appropriate resources. Thus, following Zald and Berger (1978), we expect schisms to occur least among denominations with episcopal polities, and most among congregational polities. But it is also possible to suggest a directly counterposed hypothesis. Formally centralized denominations are probably more rigid and less responsive to challenges by lower-level participants. In Hirschman's (1970) terms, failure to attend to these challenges will reduce the effective "voice" of members, and may encourage movements of "exit," increasing the tendency to schism.

The second governance variable indicates whether clergy are appointed at the level of the synod or local parish. Local clergy can play an important role in encouraging or discouraging insurgent movements, and we suspect that centralization of clergy appointments limits clerical autonomy in important ways. Where clergy are dependent on a centralized hierarchy, they may be censured for dissident activity and, in extreme cases, removed from their posts. We therefore expect that denominations that appoint clergy at the synod level will be more immune to schisms than those in which clergy are appointed at the parish level. An opposite effect is also possible, however. Like noncommissioned officers in the military, clergy provide a key transmission-point for hierarchical authority. They may enhance the flexibility and adaptability of the denomination more by being responsive to the demands of their

---

[3] Polity type is related to, but not the same as, denominational family. In these data, most denominations (73 percent), including all Baptists and Lutherans, have congregational polities. Methodists have either congregational or episcopal polities, and Presbyterian denominations have either congregational or presbyterian polities.

parishioners—and by representing those demands to superiors—than by imposing a rigid conformity from above. If this were the case, centralization of clergy appointments would encourage schisms.

Our third indicator of denominational governance reflects control over ideological resources. We examine whether the existence of a denominational seminary affects the propensity to schism. Seminaries may influence schisms in a number of ways: by providing an official source of doctrinal orthodoxy, they are likely to inhibit the emergence of competing doctrines; as centralized instruments of socialization, they may ensure the loyalty of clergy throughout their careers. Thus, the logic of the centralization argument suggests that the existence of a seminary will discourage schisms. But again a counterargument also suggests itself: the institutionalization of religious knowledge in the form of a seminary may only make incipient doctrinal disputes more visible and give contending parties an important symbolic focus for their efforts. Under this scenario, seminaries would make schisms more likely.

The second set of variables explores resource linkages between individual denominations and the wider set of religious organizations. We explore whether membership in a denominational federation affects the tendency to schism. This hypothesis borrows from Astley and Fombrun's (1983) "collective-adaptation" model of organizational survival. If we see schisms as outcomes of organizational failure, this model suggests that federation membership may help avert schism by enhancing denominational legitimacy and by giving the denomination access to a broader range of information about, for example, threatening shifts in the broader political and social environments. Also, it may give the denomination some measure of control over those environments, as when denominations lobby for preferential tax treatment or educational vouchers for sectarian schools. In our analysis, we complicate this basic argument in two ways. First, it again seems likely that federation membership could have reverse effects: when associations like the National Council of Churches adopt controversial positions on political issues, individual denominations are implicated and minorities of participants may choose to secede rather than accede to those positions. Indeed, federation membership

itself may become an issue for dispute. Second, membership in different federations may have different substantive effects. Thus, we have coded separately whether a denomination is a member of a generally liberal federation, such as the Federal Council of Churches or its successor, the National Council of Churches, or of a conservative federation, such as the American Council of Christian Churches or the National Association of Evangelicals.

Last, we explore the effects of two general organizational variables: age and size. Following Stinchcombe (1965), we expect denominations to be more vulnerable to schisms in their early years, before they have built up a secure identity and a binding culture. This "liability of newness" hypothesis was originally formulated to explain organization deaths, but it seems logical to extend it to the social hemorrhages represented by schisms as well. We use two indicators, number of churches and number of members, as measures of size. Like other measures we use, size lends itself to bivalent hypotheses. On the one hand, size is an indicator of the gross resources available to the denomination. On the other, it reflects geographic dispersion and social breadth and, perhaps, a higher degree of formalization. We include it mainly as a control variable, but we anticipate that larger denominations, because they are likely to contain more diversity and are likely to be more rigid and bureaucratic, are probably less amenable to centralized control and more vulnerable to schisms.

Data

Our sample includes all denominations that (a) existed in the United States at any time between 1890 and 1980; (b) were listed as members of either the Baptist, Lutheran, Methodist, or Presbyterian/Reformed families in the data sources described below; and (c) had at least one thousand members at any time in their history. We have excluded extremely small and unaffiliated denominations, such as the Pentecostal sects, because their life-histories are poorly recorded in available reference works. The resulting sample includes 175 denominations, distributed in the following way: 55 Baptists, 50 Lutherans, 34 Methodists, and 36 Presbyterians. Among these denominations there were 55 schisms in the period we examine. Most

denominations (79 percent) experienced no schisms, and one experienced four; the rest fell in between.

The major sources used to identify and collect data on denominations are Melton's (1978) *Encyclopedia of American Religions* and Piepkorn's (1978) *Profiles in Belief*. We supplemented these sources and cross-checked the validity of our data by referring to other reference works, including Mead (1980), the National Council of the Churches of Christ *Yearbook* (1916–84), Geisendorfer (1983), and various census reports of religious bodies (U.S. Bureau of the Census 1890, 1906, 1916, 1926, 1936). Broyles and Fernandez (1984) provided data on the founding dates of seminaries. In a few cases, we were able to gather fugitive bits of data by calling denominational officials directly. The resulting data set contains almost no missing data. The only information we were not able to gather concerns denominational size for a few denominations in the very earliest part of the study. For reasons we describe below, we are convinced that the lack of these data does not affect our findings in any important way.

The unit of analysis in this study is the individual denomination, and the data were organized in the form of event histories (Tuma and Hannan 1984). Event histories are quantitative descriptions of the life-path of each denomination as it develops over time. The primary advantage of a dynamic approach to these data, as opposed to a more conventional static approach, is that it allows us to use all the available information on the frequency and timing of events that occur to sample members and thus provides a more realistic picture of the process of schism. A second advantage to treating these data as event histories is that it allows us to code changes in independent variables as they occur over time. In this study, denominational family, polity type, centralization of clergy appointments, and federation memberships are coded as fixed attributes of denominations, while the establishment of a seminary, number of churches, and number of members (along with, of course, the occurrence of schisms) are coded as attributes that change over time. Thus, the data contain different numbers of observations for each case. For short-lived denominations experiencing no schisms, there may be only one observation; for longer-lived denominations with many schisms there are as many as eight

observations. In the end, our data set (referenced as Sutton, Liebman, and Wuthnow 1987) contains 559 observations. We describe our modeling techniques in the next section.

EMPIRICAL RESULTS

Descriptive Analysis: Rates of Schism by Family

We begin by presenting a descriptive analysis of rates of schism over time and by family. Our goals are to provide a general historical picture of the process of schism, to explore the null hypothesis that schisms vary systematically by denominational family, and to lay a conceptual foundation for the parametric analysis to follow.

The most convenient way to describe change over time in members of a sample is in terms of transition probabilities, i.e., the probability at any point in time that an individual will undergo a transition—in this case, that a denomination will experience a schism. Among the many strategies for estimating transition probabilities, the most commonly used and easily interpreted is the cumulative survivor function. Figure 1 is a plot of the occurrence of schisms, in the form of cumulative survival probabilities, broken down by family and arrayed over historical time. In this case, the stepwise decline in the lines shows the probability, at any point in time, that a denomination in any given family will survive into the next year without experiencing a schism.[4] In substantive terms, Figure 1 shows that Presbyterian denominations had somewhat higher probabilities of survival—hence, somewhat lower probabilities of schism—than denominations in other families throughout the period of the study. Methodist, Baptist, and Lutheran survival probabilities cross over each other at many points. However, this difference is probably illusory. We calculated 95 percent confidence intervals for these probability estimates. and, while we do not show them in Figure 1 for the sake of clarity, the survival probabilities for Presbyterian denominations are not significantly different from those for the other three families. In short, this graphic provides

---

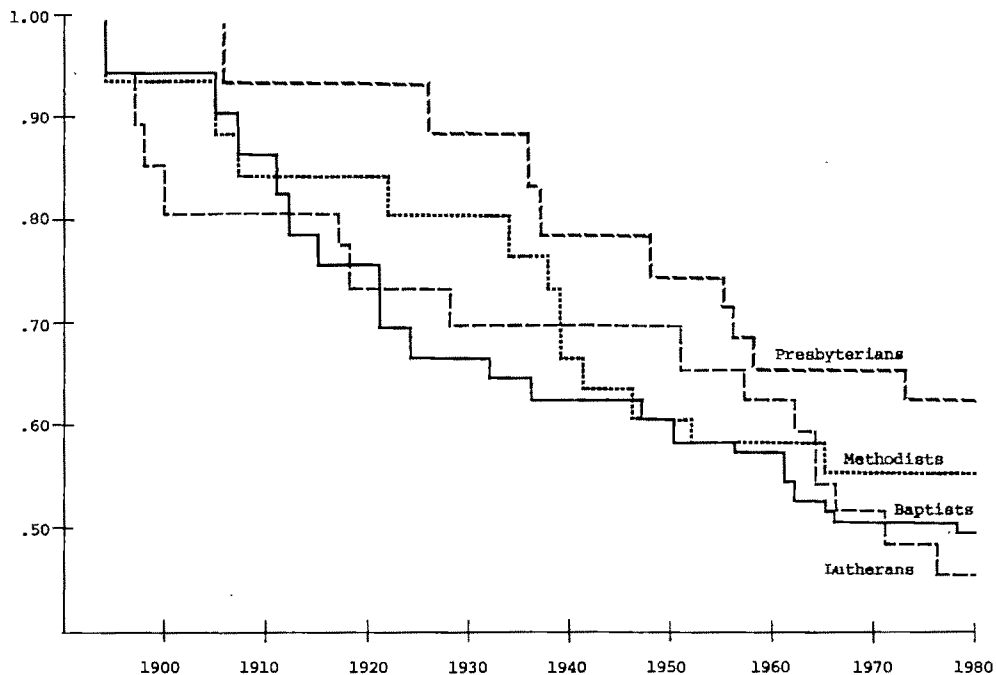[4] The plots shown here are derived from standard Kaplan-Meier estimates.

Fig. 1. Cumulative Survival Probabilities: Schisms by Denominational Family

preliminary evidence that there is no systematic variation across families in rates of schism.

If all of our exogenous variables were measured in categorical form, we could conduct much of our analysis just by constructing and examining survivor plots. But we are also concerned with continuous variables such as size, which are not amenable to graphic analysis. More important, we hope to build multivariate models that cannot be represented in a two-dimensional space.

Parametric Models of Schism

At this point we introduce a broader range of independent variables and present parameter estimates of their effects on rates of schism. But first it is important to define our dependent variable. In the subsequent discussion, when we refer to "rates of schism," or simply "rates," what we mean in formal terms is the instantaneous rate of transition from state (1), not having a schism, to state (2), having a schism. Transition rates can be defined mathematically in different ways (Tuma and Hannan 1984, pp. 71–75), but for

our purposes it is convenient to note that a transition rate is directly related to the continuous slope of the transition probabilities.[5] To use an analogy from physics, transition rates are to transition probabilities as acceleration is to velocity: they describe change in the rate of change.

In dynamic models, the null hypothesis is that transition rates are constant across individuals and over time. Available maximum-likelihood techniques (Tuma 1980; Tuma and DMA Corporation 1985) allow us to test whether denominations vary systematically in their rates of schism and derive parametric estimates of the effects of exogenous variables. Models are estimated in a linear form and can be read in a fashion similar to logit or probit regressions: b-coefficients give the multiplier effect of a unit change on the log of the rate of schism; F-tests give measures of significance for partial associations; likelihood ratio tests are used to compare nested models; and X^2 is the measure of goodness-of-fit.

---

[5] Stated more formally, the transition rate is the negative of the slope of the log cumulative survivor function.

Our first step is to estimate the bivariate effects of all the exogenous variables. Table 1 shows only the b-coefficients associated with each variable, omitting constants and X^2 statistics for convenience. To test family effects, we treated membership in each denominational family as a 1–0 dummy variable and entered three of them together in the same model (the Presbyterian dummy is excluded to avoid overidentification). The results are parameterized expressions of the plots in Figure 1. As expected, none of the families show significant variation in rates of schism. Polity type, the first of our centralization variables, is treated in the same way; here the episcopal dummy is omitted. Again we find no significant differences, and the coefficients do not show the anticipated ordering in rates: denominations with presbyterian polities, which we expected to be midway between congregational and episcopal polities, appear to be the least prone to schism.

The remaining polity variables, centralization of clergy appointments and existence of a seminary, show significant effects, but in opposite directions. In denominations where clergy are appointed at the ·synod level, the rate of schism is significantly lower than in those where congregations control appointments. The seminary variable is coded as a dummy variable in which 1 signifies the existence of a seminary. The large and significantly positive coefficient signifies that

Table 1. Bivariate Effects on Rates of Schism

Variable	Parameter	n
Family		
Baptist	.328	559
Lutheran	.454	
Methodist	.188	
Presbyterian (omitted)		
Polity type		
Congregational	.068	559
Presbyterian	−.541	
Episcopal (omitted)		
Centralization of clergy appointments	−.515*	559
Seminary	1.282***	559
Federation membership		
Liberal	.633**	559
Conservative	.746**	
None (omitted)		
Age	−.000	559
Log members (in millions)	.424***	492
Log churches (in thousands)	.467***	475

* $p < .05$.
** $p < .01$.
*** $p < .001$.

denominations with seminaries are much more likely to experience schism.

Federation membership is also coded as a pair of dummy variables; here the omitted category is those denominations that have no external federation links. Both liberal and conservative dummies have significant and convergent effects: contrary to the "collective adaptation" hypothesis, it appears that membership in any federation increases the rate of schism.

We also report tests of simple demographic factors, age and size. We tested for age effects by estimating a Gompertz model hypothesizing that rates of schism are an exponentially declining function of age. To our surprise, the resulting coefficient shows no age effect whatever. Both size variables were logged to control the effects of a few extremely large denominations. Parameters show that, regardless of whether size is measured in terms of the number of church organizations or the number of members, larger denominations are significantly more prone to schism. Note that in tests of size effects, the number of observations is reduced because of missing data. To check whether this biased our results, we re-estimated all the previous bivariate models using a reduced data set. ($n = 475$). There were no substantive changes in any of the coefficients.

The second step in the analysis was to re-estimate the effects of the centralization variables, federation membership, and age, this time controlling for size. This seemed logical because size is the most general indicator of organizational resources, and our size variables showed some of the most powerful bivariate effects. It is likely therefore that some of the other exogenous effects are spurious byproducts of denominational size; on the other hand it is possible that controlling for size will reveal latent effects. Tests showed that indicators of the number of churches and membership performed almost identically in parallel equations. We chose to use membership as our indicator of size because it allows us to use the maximum number of observations. Results from the analysis controlling for size are shown in Table 2; again constants and summary statistics are omitted.

The table reveals some interesting shifts. In all equations, the log membership variable remains positive and significant. Again, neither the family nor polity variables show

significant effects. The centralization of
clergy appointments continues to show a
significant negative effect. Here the coeffi-
cient is slightly larger in magnitude than in
the bivariate analysis. One important differ-
ence is the parameter associated with the
seminary variable. When membership is
controlled, the estimated effect is cut by 50
percent and the coefficient is no longer
significant. This suggests that seminaries are
mainly established by larger denominations,
and that by themselves seminaries have no
direct effect on rates of schism.

A second major shift involves the federa-
tion variables. The parameter for liberal
federation membership changes its sign,
becoming negative and significant, and the
effect of conservative federation membership
drops away entirely. This rather puzzling
result suggests that federation membership is
related to denominational size in some way
that confounds its direct effects. Crosstabula-
tions show that, indeed, liberal federation
members tended to be larger than other
denominations, but conservative federation
members are neither larger nor smaller than
average. We conclude that the positive
bivariate effect of liberal federation member-
ship is a spurious result of size. The negative
effect that appears when size is controlled
shows that, among large denominations,

Table 3. Effects of Centralized Clergy Appointments,
Liberal Federation Membership, and Log
Membership on Rates of Schism (standard
errors in parentheses, $n = 492$)

Variable	1.	2.	3.
Intercept	−8.016***	−8.203***	−7.906***
	(.544)	(.528)	(.528)
Centralization of Clergy Appointments	−.450		
	(.362)		
Liberal Federation Membership	−.594*	−.792**	
	(.347)	(.318)	
Log members	.488***	.505***	.424***
	(.069)	(.068)	(.064)
Chi-square	55.46***	53.85***	47.64***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

liberal federation members are significantly
less prone to schism.

Results to this point suggest three poten-
tially independent effects on rates of schism:
centralization of clergy appointments, mem-
bership in a liberal denominational federation,
and membership size. The last step in the
analysis is to combine all of these factors in a
single model and test whether their effects
persist. Complete results are displayed in
Table 3.

Column 1 shows estimates from the full
model containing all three independent vari-
ables. We note, first, that the coefficient for
the clergy variable, while still negative, is
weaker than in previous models and no longer
significant. The effect of liberal federation
membership remains negative and significant,
and that for membership is positive and
significant. In the model shown in column 2,
the clergy variable is dropped. Its elimination
slightly sharpens the estimated effects of the
other two variables, and a likelihood-ratio test
shows that the small drop in X^2 is not
significant. The model in column 3 goes one
step farther by dropping the liberal federation
dummy variable, and this omission leads to a
significant (at $p < .001$) loss of goodness-of-
fit. We conclude that model 2, which includes
liberal federation membership and log mem-
bership size as predictor variables, offers the
best available account of the process of
schism.

Table 2. Effects of Independent Variables on Rates of
Schism, Controlling for Log Membership
($n = 492$)

Variable	Parameter	Log Member-ship
Family		
Baptist	.185	.428***
Lutheran	.488	
Methodist	.107	
Presbyterian (omitted)		
Polity type		
Congregational	.438	.428***
Presbyterian	−.098	
Episcopal (omitted)		
Centralization of clergy appointments	−.688*	.432***
Seminary	.684	.369***
Federation membership		
Liberal	−.800*	.507***
Conservative	−.016	
None (omitted)		
Age	−.002	.426***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

SUMMARY AND DISCUSSION

We interpret these results in the following
way. First, contrary to conventional expecta-

tions, denominational families do not vary systematically in their vulnerability to schism. Second, contrary to the expectations generated by resource-mobilization theory, none of our measures of denominational centralization showed significant and durable effects. In simple models, the centralization of clergy appointments seemed to discourage schisms and the existence of a seminary seemed to encourage them. These relationships disappeared when other factors were controlled. Denominational age showed no association with rates of schism. The most powerful single predictor of schism is size as measured by denominational membership: the larger the denomination, the greater the tendency to schism. This is an intuitively sensible finding, but not one that offers a great deal of theoretical clarity. The membership variable stands as a summary measure of organizational resources, but it also suggests a number of organizational attributes, such as social heterogeneity, geographical dispersion, and a potential weakening of religious commitment, whose effects cannot at this point be distinguished empirically. Our best speculation is that growth raises problems of boundary-maintenance for denominations and opens opportunities for insurgent groups to appropriate resources and strike out on their own.

Finally, perhaps the most interesting finding is that, when the effect of size is controlled, linkage to a liberal denominational federation inhibits schism. This finding also suggests multiple interpretations. On the one hand, only denominations that are themselves oriented toward liberal theological and social positions are likely to join a liberal federation. This implies that doctrine after all may influence schisms and, more specifically, that denominations that permit a great deal of ideological slack may be better able to contain and co-opt dissent than more rigid denominations. On the other hand, federation linkage may operate independently of doctrine to confer a broader sense of legitimacy and collective significance on the member denomination. We suspect that both factors operate additively, that is, that avowedly liberal denominations more-or-less self-consciously seek external legitimation through federation membership. However the process works, it appears that membership in a liberal federation effectively counteracts the centrifugal effects of size.

This paper is only a first step in the analysis

of organizational vulnerability to schism. Our results suggest two paths for subsequent research. First, we plan to return to the issue of centralization and explore potential interaction effects. We suspect, for example, that polity structure may have different effects in large and small denominations. Second, we hope to add new data on a wider range of ecological factors, such as wars and economic cycles, that may influence levels of religious enthusiasm and factional disputes. Results of these analyses will yield important insights not only about religious denominations, but also about other important organizational forms, such as political parties and social movements, where diversity is produced by insurgent rebellion leading to schism.

REFERENCES

Astley, W. Graham and Charles J. Fombrun. 1983. "Collective Strategy: Social Ecology of Organizational Environments." *Academy of Management Review* 8:576–87.

Broyles, Susan G. and R. M. Fernandez. 1984. *Education Directory: Colleges and Universities 1983–84*. Washington, DC: National Center for Education Statistics.

Dornbusch, Sanford M. and Roger D. Irle. 1959. "The Failure of Presbyterian Union." *American Journal of Sociology* 64:352–55.

Gamson, William. 1975. *The Strategy of Protest*. Homewood, IL: Dorsey.

Geisendorfer, James V., ed. 1983. *Religion in America: A Directory*. Leiden, The Netherlands: E. J. Brill.

Harrison, Michael I. and John K. Maniha. 1978. "Dynamics of Dissenting Movements Within Established Organizations: Two Cases and a Theoretical Interpretation." *Journal of the Scientific Study of Religion* 17:207–24.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty*. Cambridge, MA: Harvard University Press.

Mead, Frank S. 1980. *Handbook of Denominations in the United States*. Nashville, TN: Abingdon Press.

Melton, J. Gordon. 1978. *The Encyclopedia of American Religions*. Wilmington, NC: Mc-Grath.

National Council of the Churches of Christ (Federal Council of Churches). 1916–84. *Yearbook of American and Canadian Churches*. Nashville, TN: Abingdon Press.

Niebuhr, H. Richard. 1929. *The Social Sources of Denominationlism*. New York: Henry Holt.

Piepkorn, Arthur C. 1978. *Profiles in Belief: The Religious Bodies of the United States and Canada*. San Francisco: Harper and Row.

Stark, Rodney and William Sims Bainbridge. 1979. "Of Churches, Sects, and Cults: Preliminary Concepts for a Theory of Religious Movements." *Journal of the Scientific Study of Religion* 18:117–33.

———. 1985. *The Future of Religion: Secularization, Revival, and Cult Formation.* Berkeley: University of California Press.

Steed, Mary Lou. 1986. "Church Schism and Secession: A Necessary Sequence?" *Review of Religious Research* 27:344–55.

Stinchcombe, Arthur. 1965. "Social Structure and Organizations." Pp. 142–93 in *Handbook of Organizations,* edited by James G. March. Chicago: Rand-McNally.

Sutton, John R., Robert C. Liebman, and Robert Wuthnow. 1987. DENOM8: Life Histories of American Protestant Denominations, 1890–1980 [MRDF]. Edition 8. SAS data file.

Tubeville, Gus. 1949. "Religious Schism in the Methodist Church: A Sociological Analysis of the Pine Grove Case." *Rural Sociology* 14:29–39.

Tuma, Nancy Brandon. 1980. *Invoking RATE.* Stanford, CA: Stanford University, Department of Sociology.

Tuma, Nancy Brandon and DMA Corporation. 1985. *RATE Version 2.2A.* Palo Alto, CA: DMA Corporation.

Tuma, Nancy Brandon and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods.* Orlando, FL: Academic Press.

U.S. Bureau of the Census. 1890. *Compendium of the Eleventh Census.* Washington, DC: U.S. Government Printing Office.

———. 1906, 1916, 1926, 1936. *Census of Religious Bodies.* Washington, DC: U.S. Government Printing Office.

Vrga, Djuro and Frank J. Fahey. 1970. "The Relationship of Religious Practices and Belief to Schism." *Sociological Analysis* 31:46–55.

Wilson, John. 1971. "The Sociology of Schism." Pp. 1–20 in *A Sociological Yearbook of Religion in Britain,* vol. 4, edited by M. Hill. London: SCM.

Wuthnow, Robert. 1986. "Religious Movements and Counter-Movements in North America." Pp. 1–28 in *New Religious Movements and Rapid Social Change,* edited by James A. Beckford. Beverly Hills, CA: Sage.

———. 1988. *The Restructuring of American Religion: Society and Faith Since World War II.* Princeton, NJ: Princeton University Press.

Zald, Meyer and Michael A. Berger. 1978. "Social Movements in Organizations: Coup d'Etat, Insurgency, and Mass Movements." *American Journal of Sociology* 83:823–61.

ETHNIC HEGEMONY AND THE JAPANESE OF CALIFORNIA*

ROBERT M. JIOBU

The Ohio State University

This research proposes a model to explain how ethnic minorities establish an economic niche in the host society. The model emphasizes infrastructure and proposes the concept of ethnic hegemony a situation in which an ethnic group achieves economic control over an important economic arena that interfaces with the majority. By a priori specifying the model, it is possible to see how the ethnicity of a specific group might (or might not) reinforce the infrastructure. Japanese-Americans of California were used to assess the model empirically. Since the Japanese achieved remarkable upward mobility in the face of extreme discrimination, they constitute an important test of the model. Data drawn from various historical sources show that California Japanese hegemonized a specific arena of produce agriculture, from farm labor to production and through distribution. In a more limited way, they also hegemonized contract gardening. The model of ethnic hegemony was contrasted to the usual explanation for Japanese-American success—education. Distinctions between the model and the model of ethnic enclaves are discussed. In the future, the ethnic-hegemony model might be extended and specified to other ethnic minorities.

This research proposes a model that explains how an ethnic group establishes a secure economic foothold in the broader society and builds on the assumption that, once established, the foothold facilitates the upward socioeconomic mobility of the group. While scholars agree that a group's ethnicity helps it to secure the foothold, they have not specified in detail the conditions, mechanisms, and processes. The proposed model is such a specification along infrastructural lines.

In the history of United States race relations, very few nonwhite groups have obtained socioeconomic parity with the white majority; the Japanese-Americans of California are a dramatic case in point.[1] They possessed

[1] Some points of terminology: if the ethnic group achieves socioeconomic standing commensurate with the majority, that will be termed success. If the group becomes like the majority culturally and/or structurally, that will be termed assimilation. It is possible for an ethnic group to assimilate and be unsuccessful. That would occur if the minority came to resemble the low socioeconomic segment of the majority. Original immigrants are called *Issei*. The children of the Issei are called *Nisei*, and their children are called *Sansei*. There is now a fourth generation, *Yonsei*. Sometimes *Nisei* is used to refer to

the manifest attributes of an ethnic minority— physiological visibility and cultural distinctiveness—and according to some, they were the target of more discrimination than any other group in the 20th century (Petersen 1971, Chapter 5). Yet, as many have pointed out, the California Japanese have achieved socioeconomic parity with the white majority, a conclusion that has been documented repeatedly (Varon 1967; Uhlenberg 1972; Jiobu 1976; Hirschman and Wong 1986; Jiobu forthcoming).

Data in Table 1 show that in 1980, 70 percent of California Japanese are native born and therefore may be considered an indigenous rather than an immigrant population. Whites are overwhelmingly native born, but with regard to schooling, income, and occupation, the groups are nearly identical. Moreover, multiple regression analysis of 1970 California census data showed that the Japanese equation was not significantly different from the white equation (Jiobu 1976), as the following indicates:

(1) $Y = -4{,}003 + 969\,X_1 + 522\,X_2 + 79\,X_3$ White

(2) $Y = -3{,}969 + 1088\,X_1 + 438\,X_2 + 84\,X_3$ Japanese-American

all Japanese-Americans rather than only to the second generation. This can be confusing, so more recently the term *Nikkei* is occasionally used when speaking of all Japanese-Americans.
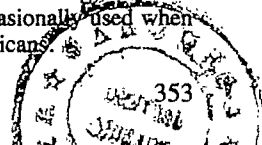
Table 1. Whites and Japanese-Americans California, 1980

Variable	Whites	Japanese-American
Percent Native Born	93.9	70.3
Mean Schooling[a]	15.1	15.4
Mean Income, All Scouces[b]	$17,216	$16,844
Modal Occupation, Category[c]	Technical	Technical

Source: Bureau of the Census 1983.

[a] Schooling: highest grade attended.

[b] Income: all sources except welfare.

[c] Modal occupation: technical, sales and administrative support.

where Y is income in 1969, X_1 is an index of occupational status, X_2 is years of schooling completed, and X_3 is age in years. The above equations were not significantly different from each other, as confirmed by the F-test. To account for possible nonlinearity in the relationship, a logarithmic transformation of earnings was also tested (results not shown), but it did not alter the conclusions. This model was replicated on 1980 Census data for California, with earnings measured in both dollars and in logarithms, and with work experience measured in years and in years squared (Jiobu forthcoming). All of these data indicate that the California Japanese have achieved extraordinary success in the face of extraordinary discrimination. As will be evident, had the California Japanese been unsuccessful, the model explicated below would not apply.

REVIEW OF THE LITERATURE

Several explanations for the failure of an ethnic group to achieve upward mobility have been proposed in the past. I review them briefly to provide the background for the model to be developed later.

Discrimination-Prejudice

Gunnar Myrdal (1944, Chapter 3) provided the classic statement of this approach. He argued that discrimination by whites prevented blacks from achieving socioeconomic mobility. The lack of mobility was then used to justify further prejudice that led to even less mobility, and so on in a vicious circle. In this argument, the onus for black failure lies with the majority; if majority discrimination

were to cease, then blacks would be able to rise socioeconomically. Myrdal was speaking of blacks, but his argument can be applied to the Japanese of California. However, because they achieved socioeconomic parity with the majority despite extreme discrimination, they seemingly contradict Myrdal's argument (see Lieberson 1980, p. 6).

Furthermore, Lieberson (1980, p. 382) claimed that the discrimination-prejudice argument misses the essential point: competition between majority and minority is the basic cause of ethnic hostility. As a result of that competition, cultural differences become salient and are interpreted as the cause (rather than result) of the hostility.

Cultural Factors

Instead of giving causal precedence to what transpires within the majority, the nature of the minority may be examined. Does an ethnic group possess a culture that facilitates success? The Japanese-American culture places a high regard on education and on the set of values contained within the Confucian ethic: hard work, sacrifice for the future, patience, and stoicism in the face of adversity (Caudill and De Vos 1956; Kitano 1976, Chapter 8; Woodrum 1985). This last value is important enough to warrant a separate name in the Japanese language: gaman.

The Japanese-American community also contains cultural devices for generating monetary capital from within the group itself (cf. Light 1972). Emigrants from the same area of Japan formed fujinkai, a club that provided recreation, mutual support, and economic aid. Japanese organizations of various kinds sponsored tanamoshi, an ethnically based organization for extending credit. Finally, the Japanese customarily gave monetary presents (koden) at weddings and at funerals and times of distress. Such organizational devices and values rooted in the group's culture help account for the often-noted absence of Japanese-Americans on welfare rolls and police blotters (cf. Petersen 1971, Chapter 6; Kitano 1976, Chapter 9).

Demographic-Ecological Factors

The size of the ethnic group is potentially important. One might hypothesize that small groups would do well because they require fewer resources from the majority, are more

easily absorbed into the majority economy, and constitute less of a political and cultural threat (Blalock 1967, p. 143). Even where they were most heavily concentrated, the Japanese were hardly more than two percent of the population, a fact consistent with this hypothesis. Still, other groups have been as small or smaller (for example, Filipinos and Gypsies) and have had less success. Thus, size alone does not explain upward mobility.

A related ecological factor is the location, congregation, and overrepresentation of a given group within a small, well-demarcated area. Such ghettoization has several possible consequences. Distance from places of employment has a negative impact on upward mobility, while compaction both increases the group's visibility (making it a target for discrimination) and reduces the likelihood of majority-minority contacts.

Wilson and Portes (1980) have applied a related concept to Cubans living in Florida. They started with the notions of primary and secondary labor markets and proposed that an *ethnic-enclave economy* includes elements of both. The enclave is characterized by many small ethnic businesses located within a ghettoized area, a characteristic suggesting a secondary labor market. However, Wilson and Portes contended that some ethnic workers may obtain a significant return on their human capital, a characteristic suggesting a primary labor market. They concluded that an ethnic enclave does exist (also see Portes and Stepick 1985).

Human Capital

Some groups possess or acquire human capital that is then converted into socioeconomic gains. Asians, in particular, have been noted for the acquisition of education (Hirschman and Wong 1986). Nevertheless, research has indicated that whites earn a greater monetary return on their educational capital than do nonwhites, including Asian-Americans in California (Jiobu, forthcoming).

Exogenous Factors

This category contains factors over which neither the minority nor majority has much, if any, control. The Japanese just happened to be emigrating when Japan was emerging as a major Eastern power, and the Japanese government was concerned about the treat-

ment of its emigrants (Lyman 1977, p. 123). For example, when San Francisco threatened to segregate Japanese school children, protest by Japan caused President Theodore Roosevelt to intercede and have the segregation order rescinded (see Bailey 1934). While the Japanese government succeeded in this instance, to judge by the Gentlemen's Agreement, the Alien Land Laws, the ineligibility of Japanese for citizenship, and the 1924 immigration law, efforts to safeguard Japanese emigrants had limited long-run success.

Timing is another exogenous factor. Early Japanese immigrants were entering the California labor force at a time when openings existed for them as domestic servants (schoolboys) and later as agricultural laborers. Moreover, California agriculture would soon experience unprecedented growth and prosperity, factors from which Japanese farmers and workers would benefit. A reasonable hypothesis is that because Japanese immigrants found, for a short initial period, a relatively open labor market in a soon-to-be prosperous industry, they could accumulate at least some capital that subsequently facilitated their movement into small enterprises, including farming.

Although all these explanations are relevant, the present model takes a different tack. Its strategy asks, "Once infrastructure has been specified, how do ethnic factors interact with it?" The primary causal variables are assumed to be infrastructural, while superstructural variables have a complementary role (cf. Harris 1979).

THE MODEL OF ETHNIC HEGEMONY

The present model builds on and incorporates the concepts of an internal labor market, sheltered labor market, middleman minority, and vertical economic integration.

1. An *internal labor market* is one that exists within a broader structure. For example, the civil service exists within the encompassing framework of government and provides job security, a career line, orderly promotion, systematic raises, and a stable work environment. In an important sense, an internal labor market is a *sheltered market:* it is protected from the adverse effects of economic cycles, political attacks, insecure tenure, and competition from other workers (cf. Freedman 1976). Because of these features, employment in internal labor mar-

kets is highly desired and if a given group becomes entrenched, it will not voluntarily relinquish its hold. In such situations, race may become the distinguishing characteristic along which hostilities develop, as exemplified by the difficulties blacks face in entering the highly sheltered labor market of construction. If, on the other hand, a substantial proportion of the minority can enter an internal labor market, their socioeconomic prospects are greatly enhanced.

2. Certain minorities have become *middleman minorities,* usually as wholesalers who mediate the economic transactions between retailers and producers (Blalock 1967, pp. 79–83; Bonacich and Modell 1980, Chapter 2). These middleman positions are often economically desirable and the majority usually dominates them—unless the minority has a monopoly on specialized knowledge or skills, or is willing to perform tasks that the majority is not.

3. An occupation or labor market in which there are disproportionately large numbers of a minority, such as Mexicans in stoop labor, is here called *ethnic saturation.* For Mexicans, ethnic saturation did not result in great socioeconomic gains. Saturation increases the likelihood that minority individuals will find employment, merely reflecting the economic realities of the labor market. The minority group is willing to work for low wages because it has limited choices. Nor does saturation necessarily mean that the minority will attain positions of power; relatively few Mexicans are in positions to determine the wage levels and employment of other Mexicans in the fields. This means that, unlike the favorable situation found in some internal labor markets, the minority suffers from low pay, limited upward mobility, little job security, and episodic employment (characteristics of a secondary labor market).

4. *Ethnic economic control* means that ethnics are in positions to make important economic decisions, such as hiring, firing, and business strategy. Note that this control differs from saturation and that the two are not always correlated. The Mexican case illustrates saturation without control. Few Mexicans are in positions to offer labor contracts, for example, so their influence is garnered mainly through the possibility of collectively withholding their labor, a tactic of limited value when demand for stoop labor is low or when the supply of stoop labor is

high. In contrast, few whites perform manual labor in South African gold mines, yet whites control the industry.

An important method of achieving ethnic economic control is through *vertical integration.* If an economic system is composed of production and distribution functions, vertical integration is the extensity and intensity of the linkages between the two segments. These linkages may consist of economic transactions, legal contracts, or interpersonal contacts and understandings.

5. Ethnic economic control and saturation together do not guarantee great economic gains. For example, ethnic businesses that sell exclusively ethnic products (such as foods) to ethnic members provide a source of income, but that activity cannot uplift an entire group.[2] To be upwardly mobile, the minority must hegemonize products or services that are in high demand by the majority, forcing the majority to deal with them. An *economic interface* between the majority and minority must exist. When it does exist, wealth is transferred from the majority to the minority, helping to uplift the entire minority group.

If all of these characteristics apply, *ethnic hegemony* exists. It is important for many reasons. First, within an ethnically hegemonized economic arena, anti-minority discrimination is largely absent. Second, hegemony means that the minority achieves at least partial command over its own economic destiny. The minority is not dependent on the largess of the majority, nor is it forced to take marginal positions in secondary labor markets. Third, within hegemonized firms and throughout a hegemonized economic arena, the minority finds an internal labor market open to it. Employment is steadier while an avenue of upward occupational mobility is available. Fourth, the majority must now deal with the ethnic group or do without. For example, unions wishing to organize that labor market are more likely to accept minority members; majority customers are more likely to deal with minority retailers and wholesalers; majority financial institutions are more likely to advance capital to minority businesses. Fifth, an economic arena provides additional sheltering because the arena as a

---

[2] Of course, certain ethnic individuals might become extremely successful, but not the group as a whole.

whole can better withstand economic vagaries than can any individual firm. Production lags, for instance, may hurt ethnic producers but may benefit ethnic retailers, who can raise prices accordingly. Last, within an hegemonized arena, ethnicity can be safely expressed, and, by being expressed, reinforced. Whether this is always advantageous can be debated, but in some cases it might be.

The foregoing model is "emergent": socioeconomic gains will occur without the overt or conscious political actions of individuals. Minority members will find greater opportunity, income, and security within an ethnically hegemonized economic arena and will be attracted to it. The majority will find commerce with the minority advantageous and will engage in it. By fulfilling infrastructural roles, an individual will contribute to the group's upward mobility without consciously being aware of it.

This point is particularly relevant to Japanese-Americans. They lacked charismatic leaders and did not engage in much self-conscious political action.[3] This was especially true of the Issei, who were politically disfranchised by the laws that prevented them from becoming naturalized citizens. And the Nisei, while citizens by birth, were not much involved in politics either. Yet, these people did gain a foothold in the economy, a foothold that constituted one very important (possibly essential) rung in the structure of socioeconomic standing.

THE JAPANESE OF CALIFORNIA

In brief, infrastructure conjoined with ethnicity enabled California Japanese to establish ethnic hegemony over a narrow and important sector of agriculture. This thesis is documented by various historical studies. Although the basic data are known, the ethnic-hegemony model organizes them in a new and theoretically meaningful way (cf. Bloom and Riemer 1949, pp. 92–93; Modell 1977, Chapter 5).

---

[3] There are no Japanese-American counterparts to Martin Luther King or Cesar Chavez. Indeed, the Japanese-American historical record is remarkably devoid of charismatic leadership. Hawaiian-Japanese followed a different political course and should not be directly compared to the California Japanese.

Beginning of Agricultural Hegemony

In 1884, the Japanese government legalized emigration and Japanese began arriving in California. Most hoped to earn their fortunes and return to Japan. Because they had no firm intentions of staying, they had little motivation to plant deep social roots. They wished to assimilate enough to "get by" and obtain employment, but beyond that, assimilation was largely irrelevant to their goals. Whatever their initial goals, though, most of these immigrants did not permanently return to Japan. They stayed and became the first generation of Japanese-Americans, the Issei.

The Issei quickly began moving into agriculture, perhaps because of their backgrounds. In his benchmark study of California Japanese, Ichihashi (1932, pp. 66–67) reproduced data from immigration records that showed that virtually no farmers immigrated until 1892, but in that year the percentage jumped to one-third (higher if laborers, most of whom might work on farms, are included).[4] U.S. Immigration Commission data revealed that by 1911, 40 percent of all Japanese in California were working as farm laborers (Petersen 1971, p. 28). Steiner (1917, p. 51) reported a not too disparate figure: 30 percent. Although these data are imprecise, they indicate the same general conclusion: Japanese immigrants had entered agricultural labor in relatively large numbers.

While their background as farmers in Japan may have predisposed them, the Japanese could not have moved into agriculture so rapidly were it not for the chronic labor shortages that then existed (Steiner 1917, p. 81). In part, the shortage resulted from an ironic consequence of the Gentlemen's Agreement of 1908. As intended, the agreement reduced Japanese immigration, but at a time when demands for Japanese agricultural workers was high. Later, the shortage was exacerbated by World War I.

The Issei soon developed a *Dano-san* system of worker organization. Under this system, a Japanese entrepreneur, called a *Dano-san* ("Mr. Boss"), would organize a gang of Japanese laborers and contract with farmers to work their fields. The *Dano-san* provided many services. He supervised his

---

[4] Fisherman were lumped together with farmers in the records, but probably few fishermen were among the group of early immigrants.

workers, furnished meals (invariably Japanese food, often prepared by his wife and daughters) and housing (sometimes temporary barracks rented from the farmer), dispensed wages, and kept records. The *Dano-san* also acted as the go-between when his laborers had to have contacts with the broader community (for example, medical care or legal work) and, in general, took responsibility for his workers' conduct. Because of the cultural gulfs between employer and employee and between employee and the broader community, this system proved indispensable to most white farmers. In a few cases, though, larger farms might by-pass the *Dano-san* arrangement and hire a so-called "Jap boss" to supervise Japanese laborers (Ichihashi 1932, p. 86).

Some *Dano-san* assumed the role of labor broker rather than direct supervisor. They established boarding houses for Japanese farm workers, renting them rooms and providing kitchen privileges and other amenities. Local farmers would then contact these *Dano-san* for laborers. When local jobs were not available, the *Dano-san* would, through contacts with other *Dano-san,* try to place their workers in other areas (see Sarasohn 1983, pp. 76–78).

The *Dano-san* system stabilized the labor force (Ichihashi 1932, pp. 173–74), a result that farmers desired. At the same time, the system greatly increased the bargaining power of the Japanese. The *Dano-san* could offer a work force en masse rather than piecemeal, and they were willing to use their bargaining power to further their own, and their workers', economic interests. The *Dano-san* might refuse to work for certain farmers or might threaten to take their work gangs off the fields if their demands were not met. In this sense, the *Dano-san* system functioned as an inchoate, ethnically organized union.[5]

Anti-Japanese Movement

Japanese immigration coincidentally occurred

---

[5] Consistent with the *Dano-san* system are the paternalistic relationships between supervisors and workers found in Japanese society (cf. Bennett and Ishino 1963). However, the *Dano-san* system may not be uniquely Japanese, as similar "Padrone" systems have existed among several ethnic groups (e.g., Italian-Americans; see Alba 1985, pp. 52–53).

not long after the Chinese exclusion movement, and so a precedent for anti-Asianism existed (Wilson and Hosokawa 1980, Chapters 7 and 8). This undoubtedly expedited the quick success of anti-Japanist forces. In a period of some 10 years, Japanese residents were disfranchised politically and further immigration was made illegal. Most of this anti-Japanist furor centered on agriculture.

In 1900, there were probably fewer than 50 independent Japanese farmers in California, with fewer than five acres each. Nine years later, there were 30,000 Japanese in California, of whom 6,000, or 20 percent, were independent farmers (Wilson and Hosokawa 1980, pp. 64–66). If these figures are only crudely accurate, they indicate that the number of Japanese farmers increased approximately 11,000 percent from 1900 to 1909. While anti-Japanists played on figures such as these, more rational observers bore in mind that the number of Japanese-owned farms was a minuscule percentage of all California farms and that 6,000 was hardly an overwhelming number in absolute terms. Nor were these figures consistent with a 1910 report from the California Labor Commission that showed that the total number of Japanese farms was 1,601, of which 56 percent were less than 30 acres in size. Ichihashi (1932, pp. 185, 193) reported data gathered from various yearbooks published by the Japanese-American Association.

At best, the data in Table 2 are crude but suggestive. They show that Japanese-controlled acreage steadily increased between 1900 to 1920, from approximately 5,000 acres to 460,000 acres, at which point it began declining. Although the increase was large in percentage terms, it never amounted

Table 2. Total Acreage Controlled by Japanese Farmers (Leased and Owned, California, 1900–1925)

	Acreage Controlled by Japanese Farmers
1900	4,698
1905	62,047
1906	97,452
1907	132,291
1908	154,801
1909	195,957
1910	194,741
1913	281,687
1914	300,474
1920	458,056
1922	330,053
1925	304,956

Source: Ichihashi 1932, pp. 185,193.

to very much of California's total agricultural acreage (cf. Ichihashi 1932, pp. 188–89). In general, Japanese-controlled acreage never exceeded 2 percent of the total, at maximum.

Even so, California began erecting a legal structure meant to drive the Japanese out of agriculture. In 1911, an alien land law passed one house of the California legislature, but not the other; in 1913, it became law. Under it, persons not eligible to become naturalized United States citizens, which included all immigrant Japanese, were barred from owning agricultural lands and limited to three-year leases. Any real property acquired in violation of the Alien Land Law was escheated to the state. The law also applied to holding companies of which the majority ownership was vested in persons ineligible for citizenship. The Alien Land Law was modified in 1920 to make it illegal for Japanese to lease land at all. These laws were popular; when the initiative affirming them was put before the California electorate in 1920, it passed overwhelmingly.

Effect of Land Law

The impact of the land laws is not clear. On paper, they should have forced the Japanese out of farming. In practice, that did not happen. Table 3 indicates the broad trend in farm ownership among the California Japanese. The percent of farm land owned hovered at approximately 15 percent in 1910, declined in 1914, and then actually increased. Between 1920 and 1922, the figure doubled, and by 1925 was 35 percent.

These data suggest that the land laws failed in their purpose, and, in a way, they did. Japanese farm ownership could still increase because the second generation, the Nisei, were United States citizens by birth and the land laws did not apply to them. Thus, many Issei vested legal titles in the names of the Nisei. A single Nisei might technically own

Table 3. Farm Land Tenure of Japanese-Americans, California, 1910–1925

Year	Owned	Not Owned
1910	15%	85%
1914	11	89·
1920	16	84
1922	33	67
1925	35	65

Source: Ichihashi 1932, p. 193.

the farms of several relatives and close friends.

Equally important, the land laws did not benefit all farming interests. Many white farmers could profitably sell their marginal lands to the Japanese, who, being short of capital, were interested buyers (Kawakami 1921, p. 52). The result was an increase in the price of these marginal lands, much to the profit of the white sellers (Ichihashi 1932, p. 183). Bankers and brokers who advanced Japanese farmers credit for supplies and capital also profited. Lenders often received interest rates 10 percent to 20 percent higher than the rates prevailing for whites while suffering fewer losses from defaults (Ichihashi 1932, p. 183). Consequently, many whites profited from the presence of Japanese farmers, and the land laws were only weakly enforced (Mears 1928, p. 253).

Hegemony over Produce Farming

Despite the alien land laws and the small proportion of California acreage farmed by the Japanese, they came to dominate a narrow, highly specialized band of agriculture: crops requiring intensive labor, specialized care, and capable of being grown in profitable quantities on relatively small plots of marginal lands. In the Sacramento area, for example, Japanese farmers grew 80 percent of the tomatoes supplied to canneries, 61 percent of the asparagus, and 78 percent of the spinach, but only 7 percent of peaches, pears, apricots, plums, and cherries (Board of Control 1922, p. 50). By the eve of World War II, Japanese farmers were producing over 90 percent of snap beans, strawberries, and celery, and approximately 50 percent of artichokes, cauliflower, cucumber, and tomatoes (Bloom and Riemer 1949, pp. 82–83). On the whole, it was estimated that by 1941–1942, Japanese farmers supplied 30 percent to 40 percent of all truck crops (Personal Justice Denied 1982, p. 43). Given the small size of the Japanese population (it varied from 0.1 percent to a high of 2.1 percent over the period 1890 through 1940 [Jiobu forthcoming], the Japanese were clearly accounting for a disproportionately large amount of California's produce.

These figures suggest the impact of the Japanese on California agriculture, but they do not indicate the importance of farming to the Japanese. Because they were so small a

population, this importance may be implicit, but scattered data from several sources help document this point. According to Ichihashi (1932, p. 165), in 1909 there were 30,000 Japanese farmers and farm hands in California, while the 1910 Census showed 41,356 Japanese (Thomas 1952, Table 3). If one accepts these figures as roughly accurate, and is willing to match them, then 72 percent of the Japanese in California were engaged in farming circa 1909–1910. This figure probably understates the impact of farming because one would wish to compute the percentage of the Japanese labor force in farming rather than the percentage of the Japanese population, but that figure was not available.

Kawakami (1921, p. 260) reported surveys taken by the Japanese Association of America in 1920. The findings are for California but exclude the nine southern counties. While this is a major omission, the data remain suggestive. They show that of the 21,784 Japanese enumerated (including children), 36 percent were classified as agriculturists and farm laborers. In 1930, Strong (1933, p. 101) conducted a 10 percent survey of California Japanese and reported that 38 percent of Japanese males were engaged in agricultural pursuits. He also noted that the Japanese consulate in California had reported a figure of 54 percent, but he did not indicate what accounted for the difference. Finally, Thomas (1952, Table 14) presented 1940 Census data showing that 42 percent of the Japanese labor force was engaged in agriculture (farm, farm manager, farm laborer, farm foreman, and farm unpaid family help).

The findings can be summarized as follows:

California:	Estimated Percent in Farming
1909–1910	72% of the Japanese population
1920	37% of the Japanese population
1930	38% of Japanese males
1930	54% of the Japanese population
1940	42% of Japanese labor force

If the data were available, one could tabulate nonfarming pursuits that were directly or indirectly supported by agriculture. For example, numerous small businesses relied on agriculture for all or part of their existence: supply and equipment stores, feed stores, restaurants, hotels, financial institutions, and ethnic vice (gambling and prostitution). Evidence from oral histories gathered by the Fullerton Oral History Program speak to these possibilities. Said one Japanese

respondent about the Japanese community of California before World War II, "Most of the workers were agricultural workers and the businesses were established more or less to cater to the needs of these people. It was more comfortable for these Japanese-speaking farmers to go to Japanese stores owned by Japanese people" (Fukasawa 1974, p. 5). Said another about Los Angeles' Little Tokyo, ". . . prior to the war, there was a large migratory worker group. During the winter, they'd work down in the Imperial Valley and that area. During the summer, they'd work in Central California, so they'd stop off here when going back and forth . . . They got some bad deals here, too from gamblers and whatever. I think those that stayed in hotels were migratory or transient" (Kunitsugu 1973, p. 2).

One would also wish to add to the figures the help donated by relatives, spouses, and children. In those days, surveyors seldom tallied the economic contribution of wives, children, and in-laws who were engaged in traditional family roles, yet their contribution of wives, children, and in-laws who were engaged in traditional family roles, yet their contributions would have been especially valuable for a farm family. In sum, even with the obvious procedural inconsistencies and methodological inadequacies, a substantial proportion of California Japanese was engaged directly or indirectly in some aspects of farming.

Hegemony over Produce Wholesaling

Considering the circumstances—small population, exclusion from the political process, anti-Japanese laws, interpersonal discrimination—one would hypothesize the Japanese farmers not to have an assured market for their crops. Such a situation did not develop, however, because the Japanese established hegemony over a segment of the produce-distribution system.

During the 1920s and continuing throughout the post-World War II period, most produce grown in California was jobbed through the two wholesale markets of Los Angeles. For both markets combined, Japanese comprised 17 percent of all commission merchants and 58 percent of all stall operators (small wholesalers) and, further, accounted for approximately 39 percent of the total dollar volume of business (Bloom and Riemer

1949, pp. 84, 85). The Japanese-Americans' presence was especially strong in the so-called San Pedro market.

Produce wholesaling was a middleman occupation and did not require a heavy investment in fixed capital. Facilities were minimal and the technology was not complex. Temporary storage and quick movement of goods were the prime requisites. According to Bloom and Riemer (1949, p. 88), Japanese wholesalers had an average capital investment of $13,000. Even at that time, the amount was small by business standards. Moreover, wholesalers sold on consignment and did not have large investments in fixed inventory.

Japanese wholesale merchants jobbed crops grown by Japanese farmers, forming a tightly linked chain of vertical integration along ethnic lines. Although direct measures of the linkage are not available, it is known that Japanese farmers dominated truck crops and that Japanese wholesalers specialized in jobbing those same crops. One may reasonably infer that the greatest bulk of produce handled by Japanese wholesalers was coming from Japanese farmers. Bloom and Riemer (1949, Chapter 3) estimated a figure of at least 75 percent.

Given their quasi-monopoly over certain crops, Japanese farmers might have sold their produce through white wholesalers, were it not for several factors. One was the possibility of race discrimination, or an unwillingness of whites to do business with them (cf. *Personal Justice Denied* 1982, p. 43). However, the extent to which this inhibited business is now impossible to document.

Japanese farmers and wholesalers were handling far more produce than the Japanese community could consume. They obviously had to have an outlet to retail consumers, an economic interface with the majority. Before World War II, Japanese greengrocers dominated produce retailing and, as such, secured the interface. Because produce had a fairly constant demand, white customers transacted business with Japanese retailers or did without. Even after World War, II, when greengrocery declined and large supermarkets entered the scene, the economic interface with whites was maintained. White buyers for supermarket chains had to deal with Japanese wholesalers or risk being caught "short."

Beyond minimal capital, the produce business had characteristics that, when combined with Japanese ethnicity, provided them with important economic advantages. Produce jobbing was a comparatively informal business and each party had to trust the other. This requirement stemmed from the characteristics of produce and the produce business. Because of changing weather and the inability to time the harvesting of crops precisely, the supply and demand of produce is not very predictable. Farmers, therefore, pushed for assurances that when the crop was harvested, a wholesale outlet would be available to them. Since vegetables are highly perishable, fast transportation is required. A large truckload of merchandise might arrive at the wholesaler's dock and be sold and delivered to the retailer within a few hours. Consequently, farmers and wholesalers had to cement dealings with trust rather than with complex legal documents. Moreover, the farmer was not at the point of sale and could not personally verify what transpired there; he had no choice but to trust the wholesaler. The converse was also true. The wholesaler depended on the farmer to ship him crops. In times of short supply, this could be critical to the wholesaler, who otherwise would have nothing to sell, and, hence, no commissions. Since the wholesaler was not at the farm site to verify the harvest, he had to trust the farmer to make good on his promises.

The vagaries of produce also meant that trust had to be established between wholesaler and retail buyers. When certain crops were in oversupply, the wholesaler could not sell everything that the farmer sent (the wholesaler was "stuck"). As produce is perishable, this placed the wholesaler in an extremely unfavorable position, and, to alleviate it, the wholesaler might send his retail customers more than they ordered, or produce of a different quality, and sometimes produce they did not order at all (practices called "pushing"). Retailers often accepted this in order to "help out," but in return they expected that the wholesaler would not "cut" their orders when produce was in short supply. A great deal of informal bargaining, arguing, and discussion predicated on much previous experience and trust made such arrangements possible.

Under these circumstances, the Japanese were in an especially good position to capitalize on their ethnicity, for it reinforced trust. Several ethnic factors can be cited. The most immediate was common subculture. The Issei spoke English poorly and could commu-

nicate more easily with each other than with whites. They shared a knowledge of Japanese customs, traditions, and practices, all of which facilitated interaction and contributed to the basis of trust. The Nisei, while perhaps not fluent in Japanese, were also intimately acquainted with this subculture.

Within the small Japanese community, friendship and kinship networks often overlapped with business networks. Sometimes farmers, wholesalers, and retailers were linked by kinship. Reputations were widely known and could be a source of embarrassment or pride to the entire kin. If someone was an untrustworthy business partner, that knowledge was quickly passed throughout the Japanese community and, conversely, if someone was especially trustworthy, others would be anxious to do business with him. The community thus judged and controlled the behavior of individuals (cf. Wellman and Leighton 1979). Kinship and friendship also played a role in securing jobs and, indeed, were the primary route to employment in the market (cf. Light 1972, p. 78). This also forged a bond of mutual obligation that reinforced the demands of trust.

The importance of this economic arena can be gleaned from the fact that 39 percent of all Japanese workers in Los Angeles County—the area with the greatest density of Japanese—were directly involved in farming, wholesaling, or retailing truck produce (Bloom and Riemer 1949, Table 7). Thomas's (1952, Table 17c) survey of Los Angeles Japanese indicated a figure of 38 percent; the two figures closely agree.

Japanese-Americans were also entering other businesses, some of which had characteristics that might fit the ethnic-hegemony model. Among them were flower raising, chick sexing, and, possibly, tuna fishing (see Modell 1977, Chapter 5).[6] The most notable example, however, is contract gardening. In Southern California, professional gardeners were common. They visited a customer's house, usually once a week, and cut the grass, trimmed the bushes, and took care of the plants and garden. From early on, Japanese saturated this occupation. Entering

---

[6] Chick sexing was once dominated by the Japanese to the extent that the technique of quickly differentiating between male and female chicks on a mass scale was kept, virtually as an ethnic trade "secret."

the gardening business required even less capital than wholesale produce (a truck and other fairly low-cost equipment), but it did require hard and long hours of work at a job many people eschewed and the aggressiveness to establish a route of customers. Usually the gardener's sons and sometimes his wife and daughters worked alongside when they could. At various times, from 20 percent to 50 percent of adult Japanese males in the Los Angeles area were engaged in this occupation (Bloom and Riemer 1949, pp. 117, 123; Tsuchida 1984).

Gardening differed from farming and produce wholesaling, but the differences did not prevent the Japanese from establishing an ethnic hegemony. For instance, Japanese gardeners patronized Japanese greenhouses and supply stores. They used Japanese trash haulers to cart away the clippings and other refuse collected during the day. The refuse truck would be parked at a convenient location, such as a Japanese-owned gas station, and gardeners would converge there at the end of the day. After unloading their trash, they shared social amenities, sometimes playing cards in the back room of the station. Also, an economic interface with the majority existed. The overwhelming majority of customers were white; majority wealth was injected into the minority economy.

At the beginning of World War II, Japanese-Americans were placed in concentration camps. The causes and consequences of the internment have been much analyzed and will not be discussed here except to note two brief points. First, the Japanese accounted for so much produce that some decision makers feared the internment might cripple the war effort (see *Personal Justice Denied* 1982). That the internment nevertheless took place implies that, while hegemony confers some protection to the minority, the protection is not absolute. Second, the internment would have seemingly devastated the Japanese beyond recovery. However, they did recover, in large part because of infrastructure of the economic arenas they had hegemonized. Japanese companies in the wholesale market went out of businesses, but the market was capitalized with shares, many of which were owned by Japanese dealers. Being paper wealth, the shares could be kept during the internment (Bloom and Riemer 1949, Chapter 3). Ironically, in some measure due to the alien land laws, 75 percent of Japanese

Table 4. Years of School Completed by California Nisei and Whites, 1940

Years of school completed	Nisei	White
Less than 4 years of high school	44%	43%
4 years high school	40	40
1–3 years college	9	10
4 + years college	7	7
Total	100	100

Source: Thomas 1952. Table 22; Bureau of the Census 1943a, Table 19.

farmers were leasing at the time (Okihiro and Drummond 1986, Table 3), and when the internment occurred, they could not lose the land they did not own. Finally, Japanese gardeners, like Japanese wholesale merchants, did not have extremely heavy investments in capital. Though they lost much, they managed to reestablished their routes after the war.

DISCUSSION

Human Capital

Conventional wisdom holds that Japanese-Americans achieved upward mobility by acquiring large amounts of human capital, especially higher education, and nothing in the ethnic-hegemony model is inconsistent with that possibility. Education could be a facilitating factor, one that helped the Japanese in their business endeavors and enabled then to branch out into the professions. Nevertheless, the case for education might be overstated. Thomas (1952), among others, studied the incarceration experiences of Japanese-Americans during World War II and reported the educational attainment of the Nisei (Thomas 1952, Table 22). Her data can be compared to the attainment of the California white population in 1940 (Bureau of the Census 1943a, Table 19). The distribution of Nisei education was virtually identical to that of whites (see Table 4).[7]

Education surely aided the Nisei in many ways, yet it did not enable them to find employment in the white economy. Thomas

(1952, Table 18) also reported the percentage of those incarcerated who had been either self-employed or employed by other Japanese before the war (see Table 5).[8]

While one might expect the Issei to work largely within the confines of the Japanese-American community, these data disclose that the same was also true of the Nisei. Because the percentages are so high, they indicate that very few Nisei had left the community for employment, whatever their educational attainment. Interviews with college placement officers during the late 1920s are informative (Mears 1928, pp. 198–99). One said, "Occasionally we have a call for a Chinese or a Japanese youth to do housework for room and board while attending college . . . No calls for Chinese or Japanese as teachers or engineers have ever come to us." Another said, "As far as I know there have been no calls come to this office for Japanese, either the first or second generation, in the lines of teaching, engineering, manufacturing, or business." Even years after World War II, no financial firm in San Francisco would hire a Japanese college graduate (Petersen 1971, p. 116). On the whole, therefore, it appears unlikely that education *alone* can explain how the Japanese managed to establish an economic niche in the California economy.

Necessary Percent

What proportion of a group must be lodged within an ethnically hegemonized arena to establish a base for socioeconomic mobility? The imperfect data reviewed here suggests that the value is approximately 30 percent to 40 percent, although much more research is required to verify that. While intriguing, this issue has not been studied very much and only a few broad considerations can be provided here.

The percentage obviously must be greater than zero because some group members must be in the arena; it must be less than 100 because only under rare circumstances would everyone (for example, children, the elderly, the retired, and the institutionalized) be in the labor force. Although difficult to measure,

---

[7] The Issei immigrated as adults and had little opportunity and probably little motivation to pursue education in the American school system. In 1940, the Sansei generation was not large and still too young to have finished its schooling.

[8] Because virtually the entire Japanese population was forcefully evacuated from California, the survey is tantamount to a survey of California Japanese at that time.

Table 5. Type of Employment Among Japanese 1940

	Percent Self-Employed or Employed by Other Japanese	
	San Francisco	Los Angeles
Issei	88%	92%
Nisei	92	94

Source: Thomas 1952, Table 18.

there are substantial indirect effects on the necessary percentage. Ethnic members who work in a hegemonized arena can indirectly support numerous ethnic businesses and individuals outside the arena. Members can "launch" the group by providing economic support for training, education, and the myriad comforts and supports that come with increased wealth. They can help open the door to the broader economy and break down prejudicial barriers, especially at the interface with the majority. In all these ways, hegemony promotes upward mobility even if only a plurality of a group is actively engaged within an arena.

Ethnic Hegemony and Ethnic Enclaves

The model of ethnic hegemony is both similar to and different from the theory of the ethnic enclave. Although detailed comparisons are beyond the scope of this paper, an ethnic enclave is characterized by compaction within an ethnically demarcated area of a city; an entrepreneurial class and subculture; existence of capital; an immigrant stream that forms a reliable and renewable source of cheap labor; an interface with the broader society; and ethnic economic growth that provides upward mobility for ethnic individuals and encourages them to stay within the ethnic community (cf. Portes and Bach 1985).

The hegemony model does not necessarily contradict several of these characteristics, e.g., that entrepreneurs and an entrepreneurial subculture are necessary for the growth of ethnic enterprises; that an economic interface is critical for both the long-term growth of ethnic businesses and for the upward mobility of persons within those businesses; and that in these processes, ethnicity is a binding force. Nevertheless, the ethnic-hegemony model differs in several major respects. One concerns the argument that an enclave ". . . permits utilization of their [the ethnic group's] past investments in human capital" (Wilson

and Portes 1980, p. 315). The Nisei were as educated as whites (see Table 4), but the Nisei could not convert their education into occupational mobility in the majority economy. To the extent that education facilitated upward mobility, it must have done so within the Japanese community (see Table 5). But even within hegemonized arenas, the majority of jobs (farming, gardening, "swamping") did not require much formal education.[9] In all probability, therefore, most Nisei were over-educated for their jobs, and one might question whether they obtained a commensurate return on their educational capital. They were aware of this, as interviews by the California State University, Fullerton, Oral History Program suggest: "I had many people remark to me, 'What's the use in your going to college when after you get your diploma, all you are going to do is either sell vegetables or push lawn mowers' " (Fukasawa 1974, p. 8). "If we went to college, we knew darn well that we were going to end up in a service station because that's about the best job we could get as a college graduate" (Yamada 1973, p. 25). Given the low rate of return on their educational capital, a sociologically interesting question for future research is "Why did the Nisei pursue education to the extent they did?"

The experiences of the California Japanese are also inconsistent with another tenet of enclave theory: "the renewal of the enclave labor through sustained immigration" (Wilson and Portes 1980, p. 314). As already indicated, the Gentlemen's Agreement, the later modifications to prevent "picture brides" from joining their spouses, and the 1924 Law effectively ended immigration from Japan. The Japanese of California were without replenishment from abroad and, therefore, enclave theory does not directly apply to their case—at least not without modification.

Perhaps the most fundamental difference between the two models is the postulation of an enclave (or ghetto) itself (cf. Wilson and Portes 1980, p. 296). To be sure, the Japanese of California formed Little Tokyos in the cities in which they lived, but these ghettos were relatively small. For instance, the Little Tokyo of Los Angeles, which was the largest on the West Coast, consisted of

---

[9] "Swampers" worked in the wholesale produce market, loading trucks, moving crates, and performing other manual labor.

just a few blocks of shops and offices, with most Japanese residences thinly sprinkled throughout the eastern part of the city (see Modell 1977, p. 57). This ecological distribution reflected the small size of the Japanese population and the additional fact that almost half (44 percent) resided in rural areas (Bureau of the Census 1943b, Table 33). Even though not heavily concentrated in enclaves, they nevertheless established an ethnic hegemony over an economic arena that was spread over hundreds of miles of farmlands, and an explanation of their case should consider this fact. The ethnic-hegemony model does so by invoking the concept of ethnic vertical integration from production through distribution, but enclave theory would, seemingly, have difficulty accounting for this fact.

Although the models share certain features, at this point in theoretical development, the enclave model might best apply to highly ghettoized populations, such as urban Cubans. The hegemony model, in contrast, applies to ghettoized and nonghettoized groups and, in addition, specifies mechanisms through which even a tiny, geographically diffused minority can obtain economic power in the face of extreme hostility.

FUTURE RESEARCH

The model of ethnic hegemony reconciles the importance of ethnicity with the importance of infrastructure, and two contrasting hypotheses may be generated:

1. Assimilation, because it weakens the ethnic group's hold on a given economic arena, leads to a decline of ethnic hegemony. Data comparing Issei, Nisei, and Sansei have shown that, by the third generation, assimilation has progressed considerably. For instance, compared to earlier generations, the Sansei outmarry more, live in non-Japanese neighborhoods, and have numerous non-Japanese friends (Montero 1981, Chapter 4; Fugita and O'Brien 1985). Consequently, the ethnicity that previously cemented relationships will weaken (if indeed it has not already done so), and ethnic hegemony will concomitantly weaken and eventually disappear. Anecdotal evidence and common observation indicate that the Japanese presence in truck farming and wholesaling has shrunk, while the declining number of Japanese greengrocers is even more self-evident. The same is

true of other hegemonized arenas, such as gardening. If later research verifies these common observations, then the model must be further specified. Hegemony provides an embattled group security from discrimination and direct and indirect avenues of upward mobility. But over time, and assuming decreasing majority hostility, the subsequent socioeconomic security of the group enables younger members to move away from ethnic arenas into the broader economy.

2. Even without embattlement, a group might maintain its ethnicity. This possibility has not been examined extensively, yet groups voluntarily support ethnicity via language schools, ethnic cultural clubs and festivals, ethnic churches, ethnic newspapers, ethnic popular culture, and so on. This described the Japanese of California prior to World War II, and for a period after the war, but probably is less accurate today.

Applying the notion of ethnic hegemony to this situation leads to a curious consequence. To maintain an infrastructure devoted to ethnic socialization, the group must possess resources. Funds are required to initiate and maintain schools, clubs, and other activities. People must be willing to support the activities monetarily and through participation. Thus, groups with greater resources can better maintain their ethnicity and so can better maintain hegemony over an economic arena.

The implications of these two hypotheses might be applied and extended in relation to various ethnic groups. The example that comes first to mind is the Chinese in Southeast Asia (the so-called Overseas Chinese). They have both achieved upward mobility *and* maintained their ethnicity, lending support to the second hypothesis. Nevertheless, caution in generalizing their case is in order. Very often, the Overseas Chinese immigrated to underdeveloped countries and brought with them entrepreneurial skills that were lacking in the host societies (Sowell 1983, p. 25). The Chinese could fill an entrepreneurial vacuum and quickly develop a commercial stratum. That, however, was not the situation encountered by the Chinese in California or by the Japanese who followed them (regarding the California Chinese, see Saxton, 1971).

The current generation of California Chinese are beginning to rival Japanese-Americans in income and already surpass

them in education (Jiobu, forthcoming). In explaining the upward mobility of the Chinese, one might ask, "Are there economic arenas that the Chinese have hegemonized?" The Chinese restaurant industry might be one such arena, perhaps another is the Chinese import-export business with suppliers in China and Taiwan and majority customers in the United States. American Jews are noted for their upward mobility—to what extent could that upward mobility be attributed to ethnic hegemony of certain economic arenas, such as the garment industry? The ethnic hegemony model may thus become useful in determining how well an ethnic minority establishes an economic niche within a host society.

REFERENCES

Alba, Richard D. 1985. *Italian Americans: Into the Twilight of Ethnicity*. Englewood Cliffs, NJ: Prentice-Hall.

Bailey, Thomas A. 1934. *Theodore Roosevelt and the Japanese-American Crisis: An Account of the International Complications Arising from the Race Problem on the Pacific Coast*. Palo Alto, CA: Stanford University Press.

Bennett, John W. and Iwao Ishino. 1963. *Paternalism in the Japanese Economy: Anthropological Studies of Oyabun-kobun Patterns*. Westport, CT: Greenwood.

Blalock, Hubert M., Jr. 1967. *Toward a Theory of Minority-Group Relations*. New York: Wiley.

Bloom, Leonard and Ruth Riemer. 1949. *Removal and Return: The Socio-Economic Effects of the War on Japanese Americans*. Berkeley, CA: University of California Press.

Board of Control of California. 1922. *California and the Oriental: Japanese, Chinese, and Hindus*. Sacramento, CA: California State Printing Office.

Bonacich, Edna and John Modell, 1980. *The Economic Basis of Ethnic Solidarity: Small Business in the Japanese American Community*. Berkeley: University of California Press.

Bureau of the Census. 1943a. *Population, 1940*, Volume IV, Part 2. Washington, DC: U.S. Government Printing Office.

———. 1943b. *Characteristics of the Nonwhite Population by Race, 1940*. Washington, DC: U.S. Government Printing Office.

———. 1983. *Census of Population and Housing, 1980: Public Use Microdata Samples, Technical Documentation*. Washington, DC: U.S. Government Printing Office.

Caudill, William and George De Vos. 1956. "Achievement, Culture, and Personality." *American Anthropologist* 58:1102–26.

Freedman, Marcia K. 1976. *Labor Markets: Segments and Shelters*. Montclair, NJ: Allanheld, Osmun.

Fugita, Stephen S. and David J. O'Brien. 1985. "Structural Assimilation, Ethnic Group Membership, and Political Participation Among Japanese Americans: A Research Note." *Social Forces* 63:986–95.

Fukawasa, George. 1974. California State University, Fullerton, Oral History Program. Interviewed by Arthur A. Hansen.

Harris, Marvin. 1979. *Cultural Materialism: The Struggle for the Science of Culture*. New York: Vintage.

Hirschman, Charles and Morrison G. Wong. 1986. "The Extraordinary Educational Attainment of Asian-Americans: A Search for Historical Evidence and Explanations." *Social Forces* 65:1–27.

Ichihashi, Yamato. 1932. *Japanese in the United States*. New York: Arno Press and the New York Times, 1969; originally published by the Stanford University Press, 1932.

Jiobu, Robert M. 1976. "Earnings Differentials Between Whites and Ethnic Minorities: The Cases of Asian Americans, Blacks, and Chicanos." *Sociology and Social Research* 61: 24–38.

———. Forthcoming. *Ethnicity and Assimilation*. Albany, NY: State University of New York Press.

Kawakami, K. K. 1921. *The Real Japanese Question*. New York: Macmillan.

Kitano, Harry H. L. 1976. *Japanese Americans: The Evolution of a Subculture*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Kunitsugu, Kango. 1973. California State University, Fullerton, Oral History Program. Interviewed by Dave Biniasz.

Lieberson, Stanley. 1980. *A Piece of the Pie: Blacks and White Immigrants Since 1880*. Berkeley: University of California Press.

Light, Ivan H. 1972. *Ethnic Enterprises in America: Business and Welfare Among Chinese, Japanese, and Blacks*. Berkeley: University of California Press.

Lyman, Stanford M. 1977. *The Asian in North America*. Santa Barbara, CA: ABC-Clio.

Mears, Eliot Grinnell. 1928. *Resident Orientals on the American Pacific Coast: Their Legal and Economic Status*. Chicago: University of Chicago Press.

Modell, John. 1977. *The Economics and Politics of Racial Accommodation: The Japanese of Los Angeles, 1900–1942*. Urbana: University of Illinois Press.

Montero, Darrel. 1981. "The Japanese Americans: Changing Patterns of Assimilation over Three Generations." *American Sociological Review* 46:829–39.

Myrdal, Gunnar. 1944. *An American Dilemma*. New York: Harper and Row.

Nobuya, Tsuchida. 1984. *Japanese Gardeners in Southern California, 1900–1941*. Pp. 435–69 in *Labor Immigration Under Capitalism: Asian Workers in the United States Before World War II*, edited by L. Cheng, and E. Bonacich. Berkeley: University of California Press.

Okihiro, Gary Y. and David Drummond. 1986. "The Concentration Camps and Japanese Economic Losses in California Agriculture, 1900–1942." Pp. 168–76 in *Japanese Americans: From Relocation to Redress*, edited by R. Daniels, S.C. Taylor, and H.H.L. Kitano. Salt Lake City: University of Utah Press.

Personal Justice Denied: Report of Commission on the Wartime Relocation and Internment of Civilians. December, 1982. Washington, DC: U.S. Government Printing Office.

Peterson, William. 1971. *Japanese Americans: Oppression and Success*. New York: Random House.

Portes, Alejandro and Bach, Robert L. 1985. *Latin Journey: Cuban and Mexican Immigrants in the United States*. Berkeley: University of California Press.

Portes, Alejandro and Alex Stepick. 1985. "Unwelcome Immigrants: The Labor Market Experiences of 1980 (Mariel) Cubans and Haitian Refugees in South Florida." *American Sociological Review* 50:493–514.

Sarasohn, Eileen Sunada. 1983. *The Issei: Portrait of a Pioneer, An Oral History*. Palo Alto, CA: Pacific Books.

Saxton, Alexander. 1971. *The Indispensable Enemy: Labor and the Anti-Chinese Movement in California*. Berkeley: University of California Press.

Sowell, Thomas, 1983. *The Economics and Politics of Race: An International Perspective*. New York: Morrow.

Steiner, Jesse Frederick. 1917. *The Japanese Invasion: A Study in the Psychology of Inter-Racial Contacts*. Chicago: A.C. McClurg.

Strong, Edward K., Jr. 1933. *Japanese in California*. Palo Alto, CA: Stanford University Press.

Thomas, Dorothy Swaine. 1952. *The Salvage: Japanese American Evacuation and Resettlement*. Berkeley: University of California Press.

Uhlenberg, Peter. 1972. "Demographic Correlates of Group achievement: Contrasting Patterns of Mexican-Americans and Japanese-Americans." *Demography* 9:119–27.

Varon, Barbara F. 1967. "The Japanese Americans: Comparative Occupational Status, 1960 and 1950." *Demography* 4:809–19.

Wellman, Barry and Barry Leighton. 1979. "Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question." *Urban Affairs Quarterly* 14:363–90.

Wilson, Robert A. and Bill Hosokawa. 1980. *East to America: A History of the Japanese in the United States*. New York: William Morrow.

Wilson, Kenneth L. and Alejandro Portes. 1980. "Immigrant Enclaves: An Analysis of Labor Market Experiences of Cubans in Miami." *American Journal of Sociology* 86:295–319.

Woodrum, Eric. 1985. "Religion and Economics Among Japanese Americans: A Weberian Study." *Social Forces* 64:191–204.

Yamada, Takeo. 1973. California State University, Fullerton, Oral History Program. Interviewed by Pat Morgan.

INTERLOCKING DIRECTORATES AND COMMUNITIES OF INTEREST AMONG AMERICAN RAILROAD COMPANIES, 1905*

WILLIAM G. ROY   PHILLIP BONACICH

University of California, Los Angeles

The debate about the separation of ownership and control has focused primarily on internal control of corporations. This study examines the relationship between ownership and control at the interorganizational level. It investigates the relationship between the structure of interorganizational title, as indicated by proprietary communities of interest, and the structure of interorganizational control, as shown by clusters of firms that recruited their boards from the same sets of directors. Secondarily the paper investigates the structure of centrality in the interlocking directorates, for which managerial theory predicts a close relationship between size and centrality. Examining American railroads in 1905, the results show that (1) while ownership and control may or may not have been separate at the level of the individual firm, at the interorganizational level the structure of title conforms very closely to the structure of control; and (2) the structure of centrality exhibits no singular hierarchy, but distinct clusters. Moreover, major companies were not highly central in the entire network, but were instead dominant within particular clusters, suggesting that centralization does not always engender power. The paper concludes that the debate over ownership and control should focus on both organizational and interorganizational levels.

This paper addresses two issues: (1) How did the rise of the corporation affect the relationship between ownership and control at the interorganizational level? The separation of ownership and control has been examined extensively at the level of the firm, but both ownership and control structures exist at the interorganizational level as well. In the early twentieth century, ownership was in the hands of "communities of interest." We will show that the social organization of proprietary title, structured through communities of interest, closely mirrored the social organization of authoritative control, structured through interlocking directorates. While the corporation made title and control potentially separa-

ble, they were empirically united at the interorganizational level, at least through the first decade of the century. (2) Were these communities of interest combined into a centralized hierarchy or Balkanized into distinct organization sets? We find no overarching central hierarchy; however, more important companies did dominate particular communities of interest rather than the full population. No single financial "superpower" reigned.

We have chosen to look at American railroads in 1905. The railroad industry was, relative to the economic landscape, more dominant than any industry before or since. The transformative period of corporate emergence has characteristics that may be visible in later periods, and it is possible to view the corporation unadorned by later embellishments. Moreover, ownership data are vastly more accessible and less concealed than more recent periods.

TITLE, CONTROL AND POSSESSION

Social-scientific analysis on the significance and consequences of the rise of the corporation has been primarily in terms of the separation of ownership and control. Berle and Means' (1932) original thesis on the

separation of ownership and control rebutted the then-prevailing wisdom that the rise of the corporation had concentrated economic power in a few capitalist hands. Instead of concentrating wealth, Berle and Means maintained, the corporation had fundamentally changed the nature of class relationships by economically disenfranchising the capitalist class, rendering them passive recipients of corporate dividends. Berle and Means' greater insight was that the corporation, by vesting proprietary authority in a board of directors that could differ from owners, created a new form of property that, for the first time, permitted institutionalized separation of ownership and control. They argued that boards of directors had become agents of professional management, no longer connected with or accountable to the legal owners.

The corporation changed the meaning of ownership of productive property, separating what had previously been one relationship into several dimensions. Recent writers typically identify three: *title, control,* and *possession* (Giddens 1975; Poulantzas 1975; Wright 1978; Scott 1979; Zeitlin 1980; Jones 1982). Under entrepreneurial capitalism, an individual who owned a productive enterprise singularly (1) had a proprietary relation, that is, held equity in a property including the right to sell it (title); (2) held an authoritative relation, that is, commanded the purposes for which property was to be used, including decisions about how profits were to be allocated (control); and (3) ran the business from day to day and administered the process of production (possession).

Berle and Means locate the most consequential transformation of corporate capitalism within the internal organization of the firm. Most subsequent work has joined the debate on these terms, accepting the individual firm as the unit of analysis, contesting whether firms are controlled by managerial or proprietary interests (Larner 1970; Burch 1972; Zeitlin 1974; Scott 1979).

CORPORATION AND SOCIALIZED PROPERTY

The corporation not only rendered these aspects separable but also transformed the organization of social relations within which they were embedded from an individual to a social level (Zeitlin 1980). Ownership became "socialized." In classical entrepreneur-

ial capitalism, one (or a few) owner(s) individually (or with one or two others) held a proprietary relationship with one (or a few) enterprise(s). In corporate capitalism, when shares are sold to the general public and when individuals own shares in many firms, title becomes social within the corporate segment of the capitalist class. Like formal title, formal authority over enterprise can be socialized, with each firm controlled by many directors and individual directors serving on more than one board. Rather than each firm being governed by an isolated board, interlocking directorates intricately join structures of formal control into a complex interorganizational web. However, both title and control are socialized very unevenly, with certain individuals and groups of individuals holding strategically decisive positions.

Communities of interest are sets of corporations linked by common ownership, either ownership by a common outsider or by mutual ownership of each other's stock. The individual firm, which had constituted the primary operative unit in entrepreneurial or competitive capitalism, can become embedded in a larger community of interest. While decisions cannot be made and executed as authoritatively among multiple firms as within the hierarchical structure of the individual firm, interfirm coordination can be facilitated by common ownership, common leadership, and institutionalized channels of communication.

By the early twentieth century, communities of interest, organized around trusts and financial institutions, had become a viable level of proprietary organization capable of concerted social action (Newcomb 1901; Moody 1904; U.S. Senate 1906; Chandler 1977).

The socialization of ownership potentially reduces owners' private interest in single firms relative to their collective interest in the aggregated profit of the corporate sector. Owners are less inclined to compete aggressively against firms in which they or their co-owners hold property.

This principle was candidly articulated in 1901 by Jacob H. Schiff, a leading investment banker with Kuhn, Loeb & Co. of New York. Describing how communities of interest arose among American railroads, he stated:

> As I take it, this community of interest idea arose in the desire of railroads, or the owners of railroads, to protect themselves against the demoralization, and, as a consequence, depres-

sion in the values of their properties, which was brought about by antipooling legislation. . . . This state of affairs [competition and falling profits] brought about a gradual coming together of the railroad interests and induced them to buy into one another's properties. For instance, if I held stock of "A" company and you held stock in "B" company, and my shares were depressed in value because you were competing with me—each of us cutting the rates of the other—our interests would evidently be better served if you owned some of the stock in my company and I owned some of the stock in your company. In other words, we had a community of interest. That is, in simple words, the process which has been going on a large scale among the railroads. . . . (U.S. Industrial Commission 1900–1902. Vol. 9, p. 770)

If both title (through communities of interest) and control (through interlocking directorates) were socialized in the sense that both resided in classes of individuals, to what extent were title and control united or separated at this level? This is the major question we will address in this paper.

CENTRALITY AND POWER

Control of the railroad industry was extremely concentrated. The Senate's Elkins Committee established that 93 individuals were capable of controlling (that is, constituted a majority of) the boards of directors of railroad systems, companies, and their subsidiaries that accounted for 75 percent of the operated mileage, 81 percent of the gross earnings, and 87 percent of the tonnage for the entire country (U.S. Senate 1906, p. 680). Such intense concentration of control in so few hands suggests the existence of a group of central companies tying the network together. The famous Pujo Committee (U.S. House 1913) thought so. They identified a unified "money trust" with Rockefeller and Morgan components that allegedly controlled American large-scale finance including the railroad industry, the industry most closely allied with finance capital. Unlike the issue of the relationship between title and control, both the managerial and class model would agree that we should find a set of coordinating companies.

Conventional wisdom links network centrality to power. As Mizruchi concluded after reviewing the relevant literature, "[T]he evidence from both small group and interorganizational research leaves little doubt that influence as the potential for control is

strongly related to network centrality" (1982, p. 52).

Because we know that the railroad system was clustered into communities of interest, it is possible that powerful firms were central within their communities of interest but not within the entire network. We will use the term *dominance within a subgroup* to refer to the extent to which a corporation has a large number of directors interlocked to other firms in the subgroup. In this type of Balkanized structure, dominance can be very important.

DATA AND METHODS

To address these questions, we examine the railroad industry in 1905. The railroad industry, "America's First Big Business," pioneered the business corporation, played the lead role in developing the institutions of finance capital for private use, and served as a model for modern managerial strategies and structures (Chandler 1965). Until the 1890s, the stock market brokerage houses, investment banks, and other institutions of corporate capitalism primarily served the railroad and related industries. Corporate capitalism *was* the railroad. Although manufacturing eventually became integrated into the structures of corporate capitalism, the railroad industry remained the core between 1890 and 1905. For example, even after the creation of large corporations in many industries, almost all interindustry interlocking directorates were with the railroads (Roy 1983). 1905 is a convenient point to examine early, yet fully developed, corporate capitalism.[1] This year followed the explosive 1898–1904 merger movement but preceded the unsettling 1907 depression. The railroad was probably at its apex, following the competitive construction wars of the 1880s and reorganization and consolidation around the turn of the century, yet prior to the long losing battle with the highway transportation industry.

Around the turn of the century, contemporary observers categorized railroad companies into communities of interest on the basis of common knowledge about financial control. To index the structure of socialized property,

---

[1] 1905 was the final year of Bunting and Barbour's (1971) annual series of interlocking directorates and adjacent to Mizruchi's (1982) beginning point in his historical overview of American interlocking directorates.

we have used the identification of companies most frequently cited in the literature and generally considered valid,[2] that of John Moody, the founder of the *Moody's Manual* (Moody 1904).

The structure of socialized control is identified by analyzing Roy's 1905 data on interlocking directorates (see Roy 1981) with techniques developed by Bonacich (Bonacich and Domhoff 1981) and here modified. The data are names of officers and directors of railroad companies with over $25,000,000 authorized capital stock. David Bunting, using the capital stock plus bonded debt figures listed in the Manual of Statistics, supplied the names of the 25 largest railroad companies (see Bunting and Barbour 1971). All other railroad companies, along with the names of their officers (president, vice president(s), secretary, treasurer) were taken from the Manual of Statistics, yielding a total of 660 individuals in 68 companies. Of these 660 individuals, 75 percent held only one directorship in the 68 companies, 12 percent held two directorships, 6 percent held three directorships, and the remaining 7 percent held four or more directorships. Two individuals were officers or directors in 17 companies. Eight companies with no interlocks were excluded from analysis leaving 60 companies.

The technique used in this paper (Bonacich and Domhoff 1981) supposes that there are "latent classes" of directors whose distinctive membership tendencies produce the observed pattern of board interlock. We assume that all directors in the same latent class have equal probabilities of belonging to boards. Moreover, within a latent class of directors, memberships in the various boards are statistically independent, so that all particular tendencies for pairs of boards to overlap must be due to their recruiting from the same latent class or classes. The technique extracts the latent classes from the matrix of overlaps between groups much like factor analysis extracts factors from a correlation matrix. As in the case of factor analysis, there is a certain amount of arbitrariness in when one stops extracting more latent classes of directors. The criterion we have used is to stop when the

――――――
[2] See U.S. Industrial Commission 1900–1902; Newcomb 1901; Parsons 1905; U.S. Senate 1906; Campbell 1938; and Chandler 1977 for similar classifications of communities of interest.

"unexplained" overlap drops below a certain amount.

More specifically, let X be a person by group matrix in which $x_{ij} = 1$ if director i is on the board of firm j and $x_{ij} = 0$ otherwise. Then $X'X$, the matrix product of the transpose of X with X, gives the overlaps between pairs of groups with the sizes of groups on the main diagonal. Let M be a hypothetical matrix in which the columns represent boards of directors and the rows represent classes of directors: m_{ij} is the number of members of board j who belong to class i. The M matrix is estimated from the data and is used to cluster boards; all the boards that apparently recruit their members predominantly from the same latent class can be clustered together. Finally, let N be a diagonal matrix whose main diagonal terms are the sizes of the latent classes of directors.

If the assumptions of the model are correct, then the following relationship holds:

$$X'X = M'N^{-1}M + D \qquad (1)$$

The diagonal matrix D is necessary because while memberships in two different groups can be statistically independent, membership in a group with itself cannot be. Estimates of M and N are formed from the eigenvectors of $X'X$, subject to certain constraints (Bonacich and Domhoff 1981). The resulting M matrix, estimating the number of members of each latent class of directors that is on each board, is used to cluster boards into sets that tend to recruit from the same classes.

Borrowing from Freeman's "betweenness" conception of centrality (Freeman 1979; Freeman, Roeder, and Mulholland 1980), we have defined centrality within systems of overlapping groups in terms of the degree to which a group connects a heterogeneous set of individuals who might not otherwise come into contact. A noncentral group recruits from only one class of individuals, while a central group is potentially more crucial in bringing disparate sets into communication. Bonacich and Domhoff (1981) propose a measure of this type of centrality. For a group j, the centrality s_j equals 1.00 when a group contains equal numbers from all the latent classes (and all group members are estimated to belong to one of the latent classes); $s_j = 0$, its minimum value, when the group contains members from only one latent class. Let q_{ij} be the estimated number of members from group

j in latent class *i* divided by the actual size of group *j* or the estimated size of group *j*, whichever is larger. The centrality of group *j* is:

$$(2) \qquad s_j = (2p/p - 1) \sum_{j > k} q_{ij} q_{ik}$$

where *p* is the number of latent classes.

Finally, a company's economic power is operationalized as the magnitude of authorized capital stock, as listed in the Manual of Statistics.

RESULTS

Using the criteria described by Bonacich and Domhoff, six latent classes were isolated with 51.2, 38.2, 61.7, 46.8, 84.2, and 59.5 members. The estimated numbers of directors of each company belonging to each of the latent classes are given in Table 1. Table 1 is the transpose of matrix *M* in equation (1).

We then cluster boards of directors according to the predominant latent class of directors from which they apparently recruit. Companies are classified by their modal latent class if the latent class contains at least 30 percent of the directors. Table 2 cross-classifies the predominant latent class within each company's board with Moody's communities-of-interest classification of the same companies. The fit between the communities of interest (the columns of Table 2) and the clusters based on boards of directors (the rows of Table 2) is remarkably close. Each cluster of firms, and its latent class of directors, is associated with one community of interest. Of the 43 companies clearly associated with a latent class of directors and classified by Moody as well, 36 (77 percent) fall into the modal main diagonal categories. For this set of 43 companies, Goodman's tau for Table 2 is .67. Moreover, none of the companies that Moody classifies as "Independent" (of any community of interest) can clearly be associated with a latent class of directors. For the complete table with all rows and columns, Goodman's tau is .52.

This correspondence of latent classes of directors and communities of interest becomes especially clear in examining specific clusters. The first cluster of firms corresponds to the Vanderbilt community of interest. Six of the seven companies in this cluster are in the Vanderbilt community. The only excep-

tion is the Chesapeake and Ohio (C&O), of which Moody stated that, while the Vanderbilts had power over it, control was shared with the Pennsylvania system (Moody 1904). Table 1 shows that the C&O included members nearly equally from Latent Class 1 and Latent Class 4, the class associated with the Pennsylvania Railroad system.

Cluster 2 is associated with the Harriman-Kuhn-Loeb community of interest. All but one of the Harriman-Kuhn-Loeb railroads belonging to any cluster belong to Cluster 2. The only exception is the Chicago and Alton: although belonging to Cluster 3, Table 1 shows that it is strongly represented by nine members in Latent Class 2.

Cluster 3 is closely associated to the Moore community of interest. Four companies conform to this pattern, while only one deviates, the Atchinson, Topeka, and Santa Fe (known as the Santa Fe), which Moody cautioned was not fully controlled by anyone. The Santa Fe, however, is in Cluster 5, strongly associated with the Morgan community, a fact challenging the conventional depiction.

Cluster 4 corresponds to the Pennsylvania Railroad community of interest, with five companies fitting this pattern and only the Reading system deviating. While Moody argues that the Pennsylvania Railroad and the Vanderbilt communities jointly controlled the Reading, he places it in the Pennsylvania Railroad community of interest because it fell within the same geographical region. He states that the Morgan community had an interest in the Reading, but no longer controlled it. These results, however, indicate that while the Vanderbilts and Pennsylvania Railroad communities may have controlled the Reading financially, Morgan had the greatest representation on the board of directors.

Cluster 5 corresponds to the Morgan community of interest. Of the nine Morgan companies associated with a latent class, eight are recruited from Latent Class 5. Only the Chicago, Burlington, and Quincy deviates, a railroad which was part of the Northern Securities Company, which Moody lists under Harriman-Kuhn-Loeb control. The Burlington is most heavily represented by Latent Class 2, which is associated with Harriman-Kuhn-Loeb interests. Thus, while Moody reports that Morgan controlled the company financially, the board of directors

Table 1. Latent Class Memberships For Boards of Directors, 1905

		Latent Class					
Code	Name	1	2	3	4	5	6
AV	Allegheny Valley	.05	.00	.00	5.11	.00	.07
SF	Atchinson, Topeka & Santa Fe	.82	.75	1.89	.86	6.81	.00
AC	Atlantic Coast Line	.00	.05	.18	.00	.06	.01
BO	Baltimore & Ohio	.62	6.25	1.17	0.89	7.47	.00
BA	Boston & Albany	.42	.12	.00	.00	.00	.00
CP	Central Pacific	.00	4.74	.00	.14	.69	.00
CR	Central Railroad of N. J.	.58	.00	.07	.00	12.12	.18
CO	Chesapeake & Ohio	7.65	.20	.00	8.06	1.40	.00
CA	Chicago & Alton	.70	8.68	14.11	.36	.92	.00
CN	Chicago & Northwestern	15.30	3.85	.20	.00	.61	.00
CG	Chicago Great Western	.00	.04	.00	.00	.09	.00
CB	Chicago, Burlington & Quincy	.67	7.47	.00	:74	4.54	.00
CM	Chicago, Milwaukee & St. Paul	1.35	.02	.28	.00	2.97	.82
CI	Chicago, Rock Island & Pacific	.61	.00	21.23	.07	.00	.00
CS	Chicago, St. Paul, Minneapolis & Omaha	13.78	1.53	.00	.00	.00	.00
CC	Cleveland, Cincinnati, Chicago & St. Louis	16.06	.00	.00	.00	.48	:12
CX	Colorado & Southern	.00	.92	4.18	.28	1.98	.00
DN	Delaware and Hudson	2.49	5.55	1.45	.13	.53	.94
DL	Delaware, Lackawanna & Western	8.87	1.40	.00	.00	6.82	.54
DR	Denver & Rio Grande	.00	5.45	.00	.14	.42	12.31
EX	Erie	2.70	4.21	2.91	.28	20.11	.00
GH	Galveston, Harrisburg & San Antonio	.00	3.89	.01	.18	.84	.00
GN	Great Northern	.00	.93	.00	.00	1.85	.00
HV	Hocking Valley	.00	.00	.40	.30	6.85	2.16
IC	Illinois Central	.00	4.79	.01	.12	.57	.00
KF	Kansas City, Fort Scott & Memphis	.03	.00	.74	.00	.00	.05
KS	Kansas City Southern	.00	8.86	5.59	.61	.00	5.02
LS	Lake Shore & Michigan Southern	18.58	.00	.00	.00	1.61	.23
LV	Lehigh Valley	3.31	.00	.03	.03	12.85	.04
LN	Louisville & Nashville	.00	.36	.13	.00	.00	.26
MK	Missouri, Kansas & Texas	.00	.00	.00	.00	.01	.45
MP	Missouri Pacific	.35	1.19	.90	.00	.00	14.06
NE	New England	.44	.00	.15	.12	.15	.21
NC	New York Central & Hudson River	18.49	1.60	.00	.00	4.44	0.00
NL	New York, Chicago & St. Louis	15.06	.00	.00	.40	1.55	.00
NH	New York, New Haven & Hartford	7.68	.00	1.09	2.07	1.92	1.18
NO	New York, Ontario & Western	4.76	.00	.43	.45	.77	·1.05
NU	New York, Susquehanna & Western	.00	.81	1.85	.08	17.83	.90
NW	Norfolk & Western	.03	.06	.01	10.45	.05	.00
NP	Northern Pacific	.32	1.36	.00	.48	10.06	.00
NS·	Northern Securities	1.11	7.61	.00	.95	8.01	.00
ON	Oregon Railroad & Navigation	.00	9.92	.00	.00	.00	5.06
OS	Oregon Short Line	.00	11.75	.43	.00	.00	10.22
PR	Pennsylvania	.21	.00	.03	17.83	.00	.12
PC	Pittsburgh, Cincinnati, Chicago & St. Louis	.25	.00	.03	17.51	0.00	.20
PF	Pittsburgh, Fort Wayne & Chicago	.33	.08	.77	10.84	1.08	.00
RX	Reading	2.76	.00	.13	.19	10.37	.48
RO	Rock Island	.91	.55	21.67	.00	.00	.00
SL	San Pedro, Los Angeles & Salt Lake City	.00	5.55	.00	.06	.24	.00
SA	Seaboard Air Line	.00	.22	5.90	.37	2.21	.75
SX	Southern	.00	.00	.42	.56	9.98	2.54
SP .	Southern Pacific	1.94	12.71	.16	.20	.18	6.92
SL	St. Louis & San Francisco	.58	.00	17.36	.00	.00	.38
SW	St. Louis Southwestern	.00	.14	.06	.04	.80	12.04
TP	Texas & Pacific	.17	2.00	.09	.00	1.05	18.77
UP	Union Pacific	1.81	13.39	2.19	.16	.00	.35
WX	Wabash	.01	.46	.00	.02	.73	12.80
WP	Western Pacific	.52	.00	.21	.03	.39	3.17
WE	Wheeling & Lake Erie	.00	1.71	.00	.00	.15	9.80
WC	Wisconsin Central	.00	.16	.03	.00	.84	.78

Source: See text.

Table 2. Railroad Companies Classified by Moody's Communities

		Communities of Interest					
Latent Class	Vanderbilt	Harriman-Kuhn-Loeb	Moore	Pennsylvania Railroad	Morgan	Gould-Rockefeller	Independent Other
1	CN, CC, LS NC, NL, CS					DL	
2		SP, UP, KS IC, GH, ON, OS			CM		
3		CA	CI, SL RI, SA			CX	
4	CO			NW, PR, AV BO, CR			PF
5			SF	RX	LV, NP, EX, NS SX, HV, GN, NU		
6						DR, MP, SW TP, WX, WE	
None	BA	CP, SL		KF	LN, AC WP	CM, MK NH, WC, NE	NO, CG, DL

Source: See text.

Note: See Table 1 for key to company names.

represented the Harriman-Kuhn-Loeb community of interest.

The last of Moody's groups, the Gould-Rockefeller community of interest, includes six companies in Cluster 6. Within this community of interest, Moody distinguishes between those more Gould and those more Rockefeller. All the railroads in Cluster 6 are identified by Moody as Gould companies. None listed as Rockefeller companies are in Cluster 6, but are distributed among Cluster 1, Cluster 3, and unclassifiable. Thus, while the. Gould and Rockefeller financial interests may have joined forces to control these companies, these results indicate that the socialized control, as manifested in control of the boards of directors, was structured quite differently, with the Gould companies clustered together and the Rockefeller not. Apparently, the Rockefellers were using devices other than control of boards of directors to coordinate the various units of their community of interest.

Moody classifies seven railroad companies as "Major independents." All fail to fall into any cluster. Moreover, for the total 60 railroad companies, the average number of interlocks is 14.3, while the seven companies classified as independent average only 5.4 interlocks. Thus, the railroads financially outside the communities of interest also fall

outside the system of interorganizational control.

These results indicate a remarkable convergence of the structure of interorganizational title and control. The contemporaneous, consensually defined proprietary communities of interest matched computed latent classes of interlocking directorates. Deeper analysis of specific firms has indicated that apparent anomalies were typically co-controlled or had recently changed control and hence did not weaken the pattern.

If title and control were united above the level of the firm, a logical next question is whether they were centralized by a powerful group of firms, the issue to which we now turn.

We examine two issues of centrality in the system. The first is the extent to which any one set of firms recruited members from all latent classes, tying together the entire system. The second is the relationship of magnitude and reputation to centrality: To what extent were large, reputedly major railroad companies more central than small, minor companies? Since the structure of interorganizational ownership (communities of interest) corresponded so closely with that of interorganizational control (latent classes of directors), and since ownership was so heavily concentrated in a few hands, one

would expect to find a highly centralized structure of interorganizational control, that is, a singular group controlling the interlocking directorate network. Moreover, since network power is widely assumed to indicate structural power (Bonacich 1972; Fennema and Schijf 1978/79; Mizruchi 1982; Mintz and Schwartz 1985), one would expect highly central and overlapping railroad companies to be the most important railroads of their period. Our results contradict these expectations: there is no tightly organized set of central and important companies. Instead, there appears to be a decentralized structure with power residing within separate communities of interest. The most central companies are not powerful but are typically subsidiaries of two or more other companies.

Table 3 displays centrality scores for the 60 companies. The 10 most central boards of directors are marked. Although the most central, these boards have significant memberships in only two or three of the six latent classes. Although there is a greater tendency for these central boards to overlap, none come remotely near to being a bridge among all classes. The odds ratio (Feinberg 1977) for the 45 pairs of overlaps among the top 10 boards averages 7.36, while for all pairs of companies it averages only 2.55, indicating that the top 10 boards are substantially more overlapped among themselves than are average firms. Despite this difference, there is not a large absolute amount of overlap among the most central companies; the overlap among the 10 most central boards averages only 11.4 percent. Although the most central companies are more likely to overlap, the amount of overlap is not very great.

Rather than uniting the whole structure, central companies typically form bridges between pairs of latent classes. For example, the railroad company with the highest centrality score is the Baltimore and Ohio, which draws more than two estimated members from only two latent classes, corresponding to the Harriman-Kuhn-Loeb and Morgan communities of interest. The second most central company is the Northern Securities Company, founded in 1901 by J.P. Morgan, James J. Hill, and others as a giant "railroad trust" (actually a holding company) among several competing interests in the northern plains and Northwest. Killed by a successful government antitrust suit in 1904, the company legally endured through 1905 only to redistribute its

Table 3. Centrality Scores for Railroads, 1905

Key	Name	Score
AV	Allegheny Valley	.07
SF	Atchinson, Topeka & Santa Fe	.20
AC	Atlantic Coast Line	.00
BO	Baltimore & Ohio	.83*
BA	Boston & Albany	.00
CP	Central Pacific	.08
CR	Central Railroad of N. J.	.12
CO	Chesapeake & Ohio	.70*
CA	Chicago & Alton	.67*
CN	Chicago & Northwestern	.35
CB	Chicago, Burlington & Quincy	.64*
CG	Chicago Great Western	.00
CM	Chicago, Milwaukee & St. Paul	.05
CI	Chicago, Rock Island & Pacific	.07
CS	Chicago, St. Paul, Minneapolis & Omaha	.21
CC	Cleveland, Cincinnati, Chicago & St. Louis	.09
CX	Colorado & Southern	.13
DN	Delaware and Hudson	.44
DL	Delaware, Lackawanna & Western	.71*
DR	Denver & Rio Grande	.56
EX	Erie	.62*
GH	Galveston, Harrisburg & San Antonio	.39
GN	Great Northern	.09
HV	Hocking Valley	.29
IC	Illinois Central	.03
KF	Kansas City Fort Scott & Memphis	.04
KS	Kansas City Southern	.72*
LS	Lake Shore & Michigan Southern	.21
LV	Lehigh Valley	.31
MP	Missouri Pacific	.28
NE	New England	.05
NC	New York Central & Hudson River	.48
NL	New York, Chicago & St. Louis	.24
NH	New York, New Haven & Hartford	.21
NO	New York, Ontario & Western	.20
NU	New York, Susquehanna & Western	.36
NW	Norfolk & Western	.05
NP	Northern Pacific	.37
NS	Northern Securities	.72*
ON	Oregon Railroad & Navigation	.42
OS	Oregon Short Line	.61*
PR	Pennsylvania	.06
PC	Pittsburgh, Cincinnati, Chicago & St. Louis	.08
PF	Pittsburgh, Fort Wayne & Chicago	.34
RX	Reading	.47
RO	Rock Island	.15
SL	San Pedro, Los Angeles & Salt Lake City	.02
SA	Seaboard Air Line	.14
SX	Southern	.30
SP	Southern Pacific	.66*
SL	St. Louis & San Francisco	.13
SW	St. Louis Southwestern	.19
TP	Texas & Pacific	.33
UP	Union Pacific	.36
WX	Wabash	.15
WP	Western Pacific	.07
WE	Wheeling & Lake Erie	.16
WC	Wisconsin Central	.01

Source: See text.
* Indicates company is among 10 most central companies.

assets to the constituent companies. Its distribution of computed directorships looks very much like the B&O, drawing multiple members from the same two latent classes. The other central companies are more marginal. Rather than dominating, five of the ten most central railroads were subsidiaries or directly controlled by other railroad companies.

Railroad companies that rank lower on the centrality measure further confirm the pattern: centrality is a matter of bridging among clusters of companies more than a matter of hierarchical ranking within a single cluster. For example, such major railroads as the Chicago, Rock Island & Pacific (4,938 miles), Pennsylvania (1,368 miles), and the Illinois Central (2,096 miles) all fall in the bottom quartile of centrality scores. All are firmly based in one latent class. A few major roads like the Chicago, Milwaukee & St. Paul (7,029 miles) or the Great Northern (5,018 miles) score low on centrality because virtually none of their directors belongs to any latent class.

Norich (1980) points out that companies may be highly interlocked, not because a discernible proprietary interest controls them, but because they are jointly controlled. This situation is more likely in a decentralized or multiple hierarchical structure. In a centralized hierarchical structure, the powerful firms will cluster together in the center of the structure. The highly central railroad companies studied here were apparently controlled by several firms, rather than a single firm, and were rarely themselves very important. In other words, in this type of Balkanized structure, centrality is not a useful index of power because power lies within communities of interest. The failure of central firms to be large and the importance of powerful firms within latent classes points to the contingent relationship between centrality and power. Only in singular, hierarchical structures will power and centrality likely covary.

We suggest that in such Balkanized structures, "dominance" within a cluster, rather than centrality between clusters, will better predict power and importance. Rather than being central in the entire network, powerful companies will tend to predominate within single clusters. The companies claiming the most members in each latent class were uniformly important and well-known companies. The New York Central and its subsid-

iary, Lake Shore and Michigan Southern, were the two companies with the most members in Latent Class 1. The former ranked 12th of the 60 companies in centrality with a score of .48, and the latter ranked 30th with a rather low score of .21. In Latent Class 2, the most dominant companies were the two Harriman companies, the Union Pacific and the Southern Pacific. The Southern Pacific was relatively central to the overall network, ranking seventh with a centrality score of .66, linking Latent Class 2 to Latent Class 1 and Latent Class 3. The Moore-controlled Rock Island predominated in Latent Class 3, but its centrality score was a rather low .15. As might be expected, the Pennsylvania Railroad claimed the most members of Latent Class 4, which was the Pennsylvania Railroad community of interest, although the company had almost no connection to any other latent class, reflecting a centrality score of .06. Latent Class 5 was dominated by the Morgan-controlled Erie Railroad. The Erie's location in the structure is consistent with Morgan's key influence in railroad affairs (U.S. House 1913). Not only did the Erie predominate within Latent Class 5, but it was one of the most overall central major railroads, linking several of the latent classes, including at least two members from four of the six latent classes. No other company included such broad representation. Nevertheless, it did not integrate the entire network. Finally, Latent Class 6 was dominated jointly by the Missouri Pacific and Texas & Pacific. Neither was ranked in the top third of companies in centrality. The composite picture is one of clustering more than a unitary centralized structure.

If centrality indicates importance, and if high capitalization indicates a major railroad, we would expect that in a hierarchical network, centrality would be statistically associated with (logged) capitalization. The results yield no such association, but rather a simple correlation of only .16 ($t=1.21$). Central companies were not especially important in this apparently Balkanized structure. Instead, the most important companies were those that dominated the six separate clusters of firms. We isolated the three firms that had the most members in each of the six latent classes, eighteen firms in all. The correlation between a dummy variable indicating membership in this set and (logged) capitalization was a highly significant .34 ($t = 2.70$, $p <$

.01). Clearly, major companies dominated single latent classes rather than coordinated the entire structure.

What is the significance of a Balkanized structure among railroad companies? The results indicate that the term *communities of interest* more aptly describes the uneven structure of socialized ownership than does the term *hierarchy of domination*. Rather than a central focal point tying the network of interlocks together, the companies were organized into six coherent clusters.

It is important that this clustering existed within a homogeneous type of corporate organization. The property relations of corporate financing, publicly marketed securities through the New York and London exchanges, location of authority within the board of directors, and dependence on large amounts of finance capital characterized all the companies. The differences in the structural location of railroad companies in the interlocking network indicated quantitative differences within a single mode of operation rather than an economically heterogeneous set of firms. Although the communities of interest were integrated by both common financial control and common leadership, at the supra-firm level there was no structural separation of ownership and management.

CONCLUSION

This paper has examined the structure of economic ownership and authority when the American economy was being transformed from entrepreneurial capitalism to corporate capitalism. The corporation not only changed the way in which a particular enterprise was owned but also the way in which ownership was structured throughout the capitalist class. Capitalists who had previously faced each other as competitors now joined together as associates, linked by the socialization of capital and clustered into communities of interest.

First, we have demonstrated that the structure of interorganizational ownership corresponded remarkably closely with the structure of interorganizational control. At the social level, if not the individual level, the capitalist class simultaneously exercised proprietary title (or ownership) and authoritative control of the corporations. We suggest that the ownership-control debate, by focusing on the individual firm and the individual owner,

mistakenly ignored what may be the more important unit, the community of interest. Thus, whether or not Berle and Means are correct that the corporation disenfranchised the individual owner, the corporation cannot be said to have dissolved the capitalist class. On the contrary, by socializing ownership of productive property, the corporation consolidated the capitalist class by restructuring capital into a new institutional form of proprietary and authoritative relationship. To analyze the structure of property from this period onward, the social level of analysis must be as thoroughly considered as the individual firm. Contrary to some authors (Burt 1982; Pfeffer and Salancik 1978; Allen 1974) who see interlocking directorates as reflecting interorganizational managerial coordination, in the initial stage at least, the structure of interlocking directorates projected the structure of property. The relations among companies were relations among owners, not just managers. The interlocks coordinated capital investments, not just operations. They were clearly capitalist relationships.

Second, we have demonstrated that the vertical dimension of centralization in the structure of authority deviated from the structure of power conventionally considered in the industry. The railroads' polymorphous communities of interest were not organized or centralized into a single, superordinate dominating structure. Because the structure was disjointed or Balkanized, there was no relationship between firms' centrality and their reputed power or size. The central companies were not the most important or powerful firms. The most dominant railroads of the time did not tie together different communities of interest but typically recruited members from single communities of interest; power existed within the communities of interest rather than in any overall coordinating structure. The relationship between centrality and power, long taken for granted in network analysis, must be treated as contingent rather than systemic (Bonacich 1987).

While it is beyond the scope of this paper to fully explain the conditions under which centrality may coincide or depart from power, one plausible factor is the content of links in the network. The most common content for network links is communication of knowledge, which is treated in various network studies from laboratory experiments to interlocking directorates. If nodes are linked by

information flow, actors bridging between other nodes will be more capable of domination than isolates. Central firms will be powerful. However, not all types of links give incumbents equal advantages. In a bargaining situation, for example, a monopolist is strengthened through the isolation of those to whom he sells, even though their isolation may reduce his own centrality in the entire economy. Proprietary linkages may be structured less by network centrality than by the more direct domination of ownership. In any case, the disassociation of centrality and power found here compels a new and more nuanced look at how they are related in general.

Our findings that socialized title closely corresponded to socialized control and that major companies dominated rather than bridged clusters is no "smoking gun" to convict managerial theory. In addition to the corporation's profound effect on the internal organization of the firm, it engendered an equally consequential transformation of the meaning of property at the social level. Formerly, owners had competed against each other in the marketplace. The parcelization of ownership into many shares made it possible to unite proprietary interests by interlocking ownership, while the parcelization of authority into boards of directors made it possible to coordinate different firms' activities.

The most cautious conclusion to be drawn from our results is that early corporations were organizationally united by coinciding structures of ownerships and directorships, without implying anything about who wielded power within the corporation. There is no doubt that a major impact of the development of the corporation was transforming property into separable components. Title, by definition, remained with the owners. Possession, or mundane management, was by this time in the hands of professional management. At issue is the concept of control, the ability to make strategic decisions. Directors legally make these decisions. Managerialists treat directors as agents of the firm who exercise control at the behest of management. Class theory treats the firm as owners' property and directors as owners' agents. The fact that the social organization of directors coincided with the organization of ownership does not definitely prove that directors controlled firms, especially at the firm unit of analysis, but it is more consistent with the class model

than with the managerial model. The findings show how the capitalist class was consolidated rather than dispersed.

Although our data cannot determine whether our results reflect only a transitional phase between entrepreneurial and managerial capitalism, we have uncovered a fully developed corporate structure with a unified structure of socialized title and socialized control. To the extent that possession, title, and control of corporations have actually become differentiated among different actors, the explanation must be grounded in forces other than the inherent features of the corporation itself.

REFERENCES

Allen, Michael Patrick. 1974. "The Structure of Interorganizational Elite Cooptation: Interlocking Corporate Directorates." *American Sociological Review* 39:393–406.
Berle, Adolf A. and Gardiner C. Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.
Bonacich, Phillip. 1972. "Technique for Analyzing Overlapping Memberships." Pp. 176–85 in *Sociological Methodology, 1972*, edited by Herbert L. Costner. San Francisco: Jossey-bass.
———. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92:1170–82.
Bonacich, Phillip and G. William Domhoff. 1981. "Latent Classes and Group Membership." *Social Networks* 3:175–96.
Bunting, David and Jeffrey Barbour. 1971. "Interlocking Directorates in Large American Corporations, 1896–1964." *Business History Review*.
Burch, Philip H., Jr. 1972. *The Managerial Revolution Reassessed*. Lexington, MA: Heath.
Burt, Ronald S. 1982. *Toward a Structural Theory of Action: Network Models of Social Structure, Perception and Action*. New York: Academic.
Campbell, E.G. 1938. *The Reorganization of the American Railroad System, 1893–1900*. New York: Columbia University Press.
Chandler, Alfred D. Jr., ed. 1965. *The Railroads: The Nation's First Big Business*. New York: Harcourt, Brace & World.
———. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge: Belknap.
Feinberg, Stephen E. 1977. The Analysis of Cross-Classified Categorical Data. Cambridge: MIT Press.
Fennema, Meindert and Huibert Schijf. 1978/79. "Analyzing Interlocking Directorates: Theory and Methods." *Social Networks* 1:279–332.
Freeman, Linton C. 1979. "Centrality in Social

Networks: Conceptual Clarification." *Social Networks* 1:215–39.

Freeman, Linton C., D. Roeder, and R. R. Mulholland. 1980. "Centrality in Social Networks: II. Experimental Results." *Social Networks* 2:119–41.

Giddens, Anthony. 1975. *The Class Structure of the Advanced Societies*. New York: Harper and Row.

Jones, Kelvin. 1982. *Law and Economy: The Legal Representation of Corporate Capital.* New York: Academic.

Larner, Robert J. 1970. *Management Control and the Large Corporation.* Cambridge, MA: Dunellen.

Marx, Karl. [1867] 1967. *Capital: A Critique of Political Economy,* vol. III. New York: International.

Mintz, Beth and Michael Schwartz. 1985. *The Power Structure of American Business.* Chicago: University of Chicago Press.

Mizruchi, Mark S. 1982. *The American Corporate Network: 1904–1974.* Beverly Hills, CA: Sage.

Moody, John. 1904. *The Truth About the Trusts.* New York: Moody Publishing.

Newcomb, H.T. 1901. "The Recent Great Railway Combinations." *American Monthly Reviews* (August): 163–74.

Norich, Samuel. 1980. "Interlocking Directorates, the Control of Large Corporations, and Patterns of Accumulation in the Capitalist Class." Pp. 83–104 in *Classes, Class Conflict, and the State,* edited by Maurice Zeitlin. Cambridge, MA: Winthrop.

Parsons, Frank. 1905. *The Railways, the Trusts, and the People.* Philadelphia: Taylor.

Pfeffer, Jeffrey and Gerald R. Salancik. 1978. *The External Control of Organizations: A Resource Dependence Perspective.* New York: Harper and Row.

Poulantzas, N. 1975. *Political Power and Social Classes.* London: Humanities.

Roy, William G. 1981. "The Vesting of Interests and the Determinants of Political Power: Size, Network Structure, and Mobilization of American Industries, 1886–1905." *American Journal of Sociology* 86:1287–1310.

———. 1983. "The Unfolding of the Interlocking Directorate Structure of the United States." *American Sociological Review* 48:248–56.

Scott, John. 1979. *Corporations, Classes, and Capitalism.* London: Hutchinson.

U.S. House of Representatives. 1913. *Money Trust Investigation. Investigation of Financial and Monetary Conditions in the United States.* Washington, DC: Government Printing Office.

U.S. Industrial Commission, 1900–1902. *Reports.* Washington, DC: Government Printing Office.

U.S. Senate Committee on Interstate Commerce. 1906. *Regulation of Railway Rates.* Digest of the Hearings, 59th Congress, 1st Session, Document No. 244. Washington, DC: Government printing Office.

Wright, Erik O. 1978. *Class, Crisis, and the State.* London: New Left.

Zeitlin, Maurice. 1974. "Corporate Ownership and Control: The Large Corporation and the Capitalist Class." *American Journal of Sociology* 79:1073–1119.

———, ed. 1980. *Classes, Class Conflict, and the State.* Cambridge, MA: Winthrop.

PROFESSIONALISM, VARIATION, AND
ORGANIZATIONAL SURVIVAL*

DAVID L. TORRES
University of Arizona

Organizational-ecology theorists have traditionally examined the effects of institutional variables in the same manner as task-environment variables—as affecting the selection process. This paper illustrates that institutions, in this case professionalism, may also affect variation of structural forms within a niche. Using a case study of the funeral industry, important effects of the institutional environment on variation are identified. The study contributes to the ecological perspective by showing that, in some cases, the selection game is fixed because variation is not random.

Ecological views have gained considerable popularity among students of organization theory, but not without some criticisms. One is that the theoretical assumptions of population ecology work well with populations of smaller organizations with dispersed resources (Astley and Van de Ven 1983), but not as well with populations of larger organizations with some control over their environment (Betton and Dess 1987; Aldrich 1979). Another is that ecology theory tends to underplay the issue of human motives and the power of individuals or interest groups to affect organizational performance (see, for example, Perrow 1986). A final critique of ecology theorists is that they have tended to over-study the selection process, while limiting examinations of organizational variation (Hawley 1981).

This research is intended to show that the criticisms are valid in some cases and that they have important implications for population-ecology theory. Specifically, the study shows that, in the case of institutional environments, the usual focus on selection dynamics may only partially explain survival of organizational populations. Institutional effects on variation that are allowed to occur within environments must also be considered. In other words, instead of always assuming that the institutional environment (in some

abstract form) selects the best fit form, it may be fruitful to investigate the possibility that institutional forces prohibit some forms from competing before selection or that institutional forces set up definite rules of construction, preempting the process of selection. As Gould (1977, p. 42) has noted with regard to natural selection, fit can be *expressed* as differential survival, but not necessarily *defined* by it.

The argument is based on an analysis of professionalization and its effects on population dynamics of funeral establishments. The study illustrates how a concerted effort to gain control of an environment led to the formation of a homogeneous and stable population. Professionalism as an agenda is shown to have directed variation of funeral establishments quite explicitly by articulating rules for structuring organizations and outlawing structures that threatened the existence of the accepted form of funeral establishment.

ECOLOGICAL PERSPECTIVES
ON EVOLUTION

The importance of social and political forces to organizations' survival has not been neglected by students of organizational environments (Meyer 1978; Pfeffer 1978; Aldrich 1979). Meyer (1978, p. 365) views organizations ". . . as competing politically for institutional legitimacy in contemporary environments and less as competing technically for market advantages." Carroll and Huo (1986, p. 838) define the institutional environment as "external factors that indirectly affect an organization through societal norms, resources, and constraints." However, most

ecology research has examined institutional variables only as they affect selection (Carroll and Huo 1986; Singh, Tucker, and House 1986; Wiewel and Hunter 1985). Questions about how and why legitimacy is created have received less attention (but see Hirsch 1986), as has the possibility that, once having acquired legitimacy, organizations might use that legitimacy for purposes beyond those directly affecting competition.

The basic precepts adopted by population and community ecologists have been covered in other works (Carroll 1984; Astley 1985). Presented here is a synopsis of the treatment of two of three integral components of ecology models: variation and selection (Perrow 1986; Carroll 1984; Pfeffer 1982). The third component, retention, is discussed either directly or indirectly in works using the institutionalization perspective (see, for example, DiMaggio and Powell 1983; Meyer and Rowan 1977), and in other works addressing the basic question of inertia in organizations (Starbuck 1983; McKelvey and Aldrich 1983).

In most research using the population-ecology perspective, rates of founding, net mortality, or other variables most closely associated with selection are dependent variables, while variation is assumed or used as a covariate (see, for example, Delacroix and Carroll 1983; Freeman and Hannan 1983; Freeman, Carroll, and Hannon 1983; Hannan and Freeman 1984; 1987). Best summarizing the role of variation in contemporary ecology research is Carroll (1984, p. 74), who states that, although organizational variation is an essential precondition of selection, it is not necessary for the modeling of selection processes in ecology models.

By adopting this perspective, ecologists further assume that variation of organizational structures is relatively random. This has not been especially problematic for much of the early research, which examined dynamic populations like the restaurant and semiconductor industries. Random variation of structural forms is very likely to exist in these industries because much uncertainty surrounds the question of what influences survival. As research turns to institutional environments (Singh, Tucker, and House 1986; Wiewel and Hunter 1985; Carroll and Huo 1986), however, assumptions about random variation of structures become more questionable.

Environments strongly influenced by institutional forces are likely to constrain what constitutes a "proper" organization; that is, characteristics leading to increased probability for survival are known before organizations are constructed, and the structuring of organizations is guided by such knowledge. In essence, variation of structures is constrained by norms and regulations that surround the structuring process: the more knowledge about proper structuring of organizations, the more deterministic the variation. Darwin (1859, p. 90) recognized similar deterministic tendencies in selective breeding, of which he said that, since men constructed standards by which creatures or plants were to be selected, variation was accumulated in directions that men established.

In strongly institutionalized environments, the question of why variation is accumulated in certain directions and not in others may be more important than the question of which organization is "best fit" to a particular environment. For if variation is not random, selection "merely eliminates the unlucky individuals who do not vary in the appropriate way" (Gould 1977, p. 12). Few studies have looked at variation empirically. Works by Stinchcombe (1959; 1965), Hawley (1981) and Astley (1985) discuss variation at the theoretical level, although Stinchcombe (1959) does compare two industries to explore the question of what causes variation, not what constrains it.

A second assumption implicitly made as a result of emphasizing selection over variation is that incremental change, as opposed to discontinuous change, is the stuff of structural evolution. Yet, Astley (1985; see also Tushman and Anderson 1986) argues that discontinuous change can come about through social or political causes that disrupt the continuity of previous environments. These social or political forces, not necessarily tied to technological innovation, lead to the "provision of open environmental space" (p. 238). The recent opening of China to capitalism is an example.

The significance of this argument is that, in the case of discontinuous change, selection dynamics provide only a partial explanation of organizational survival. What is also important in new niches is the variation of new entrants. Given self-interest to survive and some ability to control their environment, it is very conceivable that early entrants will attempt to exclude organizational forms that follow—not necessarily by out-competing them.

Both the questions of constraining variation and of consequences of discontinuous change will be addressed in this paper. The study illustrates how one discontinuous change, gaining and losing professional status, affected survival chances for different populations of organization. The central purpose of the study is to argue that, in certain cases, current models of population ecology, with their focus on selection, need to be extended to include environments' effects on variation.

PROFESSIONALISM AS AN INSTITUTIONAL VARIABLE

At base, professionalization involves the transference of policy-making authority from the state to an occupation (Brown 1979; Starr 1982; Lubove 1980; Lynn 1965). Self-regulation is considered necessary in professions because the special expertise and training that professionals possess makes others unable to evaluate performance or determine the best policies for such occupations (Hughes 1965). Because of the relative autonomy that professions have, they have been said to hold monopolies over certain services (Larson 1977; Johnson 1972; Freidson 1970). A professed commitment to service to clients and code of ethics are personal assurances that professional powers will not be misused, while state boards of practice, comprised of colleagues, serve as formal policing and sanctioning bodies.

The power to self-regulate can affect organizational structures in two ways. First, professions may formulate and enforce policy that excludes certain structures. For example, the legal profession prohibits corporately owned law firms in many states on grounds that commitment to the code of ethics may be affected by corporations' commitment to the bottom line. Second, because professions establish standards and procedures for licensing, they may expressly forbid certain orientations or ideologies that promote the introduction of new structures. At a public hearing on professional licensing for funeral directors, for example, the state board chairman testified that a cemetery owner wanting to establish a funeral home on cemetery property found it very difficult to obtain a license — no funeral director would allow him to do an internship. Without an internship, he was unable to obtain a license (data from taped hearings). In other instances, individuals with communist tendencies have been denied licenses (Friedman 1962).

DATA AND METHODS

The funeral industry is significantly different from industries commonly examined by ecologists. The industry comprises a relatively homogeneous and stable population of establishments. This is unlike most niches studied by ecologists, which are usually environments with dynamic net mortality and variable structures. Undeniably, funeral establishments have not always enjoyed such stability. Indeed, this research intends to show that professionalization was intricately connected with the homogenizing and stabilizing effects and, further, that professionalization within the industry is leading to more varied and competitive structures. A second characteristic of funeral directing is its "semi-profession" or "marginal profession" status (Crouch 1971). Although the industry gained full professional status in most states, it never attained the legitimacy of more developed occupations like law or medicine. The status of funeral directors as professionals peaked during the 1950s. During the 1960s and early 1970s, the industry was subjected to bad publicity and investigations by federal agencies, among them the Justice Department and Federal Trade Commission. As a result, the momentum of professionalizing was lost in all states, and, in at least one state, professional authority for funeral directing was abolished. Consequently, one can examine both the effects of professionalization and declining professionalization within the same industry, more effectively isolating professionalism and its effect on structure.

Data for this research are archival (Torres 1983), involving two main types of information. The first deals with identification of structures that exist or have attempted to exist in the industrial niche. Statistical data from *Census of Service Businesses, U.S. Industrial Outlook, The Statistical Abstract of the United States,* and *Vital Statistics of the United States* were used to document the number and size of funeral establishments and crematories in existence from about 1900 to the present. To identify and approximate the number of structures that *attempted* to survive within the environment, two trade journals and conference proceedings were reviewed. *Casket and Sunnyside* and *Mortuary Management* were selected

from among other journals because of their objectivity in reporting, extent of coverage, and age. *Casket and Sunnyside* traces its beginnings to 1871, with the founding of *Casket,* which later merged with *Sunnyside* to form the current journal. Similarly, *Mortuary Management* traces its beginnings to 1914, with the founding of its forerunner, *The Pacific Coast Undertaker.* Data from these two journals were complemented with official records of the annual conference proceedings of the National Funeral Directors Association. The journals and proceedings were scanned from cover to cover for the period covering circa 1917 to 1981, which includes the period preceding professionalization and the beginning of deprofessionalization. Other archival sources included books, dissertations, reports, and articles stored at the National Foundation for Funeral Services, Evanston, Illinois. Virtually every significant movement made by new and different types of organizations was covered by the journals or association. The high visibility of new structures can be tied to the industry's interest in upgrading standards to professional levels. New forms were seen as a possible threat to this mission and their movements closely monitored.

The second focus for data gathering was on reconstructing the history of professionalization in the industry. The trade journals, conference proceedings, dissertations, and books mentioned above were used for this task as well. In addition, 50 oral, semi-structured interviews were conducted, 29 with funeral directors in a large midwestern city, and others with representatives from the Federal Trade Commission, state officials, officials from the National Selected Morticians and the National Funeral Directors Association, state and local funeral directors' associations, and suppliers of funeral-service supplies. The taped interviews provided data on personal histories of funeral directors, funeral home operations, views on the efforts to professionalize, and current status of the industry. Non-funeral directors were asked more specific questions about the role of their agencies relative to the industry. Interview data are cited as used in the text.

A BRIEF HISTORY OF PROFESSIONALIZATION IN THE FUNERAL INDUSTRY

A more complete history of professionaliza-

tion of the industry is given in Torres (1983); only important highlights will be discussed here. Beginning with the establishment of the National Funeral Directors Association in 1882, funeral directors began transforming a relatively open occupation into a more exclusive one. To rid the occupation of overly competitive and often unscrupulous undertakers, including curbstoners who plied their trade in the streets, prominent undertakers saw education and professionalization as a means for upgrading the occupation (*Mortuary Management* 1929a, p. 33; *Mortuary Management* 1929b, p. 22). The expertise allowing funeral directors to seek professional status was embalming, which tied them to surgeons, physicians, and physiologists, who often served as instructors of embalming (Habenstein and Lamers 1955, p. 471). By the 1940s, however, embalming knowledge was not progressing, except for "improved" products put out by embalming fluid manufacturers (Renouard 1944, p. 34). Efforts by some embalmers to unionize (*Casket and Sunnyside* 1941b, p. 50; *Casket and Sunnyside* 1948, p. 70–71; *Mortuary Management* 1951, p. 28; 1952, p. 40; 1953, p. 34; 1956, p. 6) prompted leaders within the industry to redefine expertise as knowledge in counseling, grief, legal requirements, and customs (interview with official from trade association). Despite the change in professed expertise, funeral directors were able to upgrade educational requirements to professional standards. In 1972, accrediting powers were granted to an industry-sponsored organization, the American Board of Funeral Service Education, by the U.S. Office of Education. This gave the funeral industry formal power to regulate education in the field (*Mortuary Management* 1972, p. 42). However, the American Board and its forerunners had been accrediting schools on an informal basis since the 1920s.

State legislative activity creating self-regulating boards of practice occurred in two waves. The first wave centered on establishing state boards to regulate embalming and disposal of dead bodies in the early 1900s. The movement was sparked by a national trend promoting sanitation and public health (Illinois Bureau of the Budget 1982, p. 2; Kansas Legislative Division of Post Audit 1980). The second wave comprised acts to establish self-regulation for funeral directors in the 1930s (see, for example, Texas Sunset

Advisory Commission 1978, p. 4). Almost every state eventually established self-regulation for the industry (Neilson and Watkins 1973, p. 88; Federal Trade Commission 1978, p. 103). Of those state boards not comprised exclusively of funeral directors, eleven included a member from the state board of health, two included other government officials, and three included a lay person. In 1978, thirty-six states gave its board exclusive control to regulate the occupation, six state boards initiated proceedings before the state regulatory authority, and eight boards acted only on filed complaints (Neilson and Watkins 1973, p. 103).

The force behind professionalizing was the National Funeral Directors Association (NFDA) and its state affiliates. The NFDA dispensed information, assistance, and funds in efforts to establish professional authority from state to state (Gilligan 1928, p. 16). During the 1940s, the NFDA also accepted the responsibility of reviewing proposed legislation in the states, in order to "avoid embarrassing court decisions" (NFDA 1940, pp. 70–71). Eventually recruiting about 75 percent of all practitioners, the NFDA was also an effective lobbying group. During a Federal Trade Commission investigation in the late 1970s and early 1980s, for example, it was heavily involved in mobilizing support "urging funeral directors to solicit anti-rule letters and testimony and to lobby their Congressmen and Senators to put pressure on the FTC" (*Mortuary Management* 1978, p. 22). The *Washington Post* (1984, p. 30) regarded this effort as "one of the longest challenges from an industry group in commission history." Funeral directors made substantial progress toward professionalism through the 1950s.

POPULATIONS AND
POPULATION DYNAMICS OF
FUNERAL ESTABLISHMENTS

The following data are intended to show that survival chances of traditional funeral establishments, which tend to be a homogeneous group, are enhanced by a lack of competition. Traditional funeral homes are small, offer fair but not lucrative returns, and are operating within a market that is naturally bounded by the number of deaths occurring per year. The industry is at or over capacity, but lack of competition within the industry allows many marginal establishments to survive.

Funeral homes in general had the lowest failure rates of any type of business in the U.S. in 1985 (*Wall Street Journal* 1985, p. 25). A failure rate of 10.9 per 10,000 for that year is still higher than that reported for other years. Only three funeral homes failed in 1974, for example, for a rate of less than 1.4 per 10,000 establishments (Herringer 1976, pp. 25–26). Only eight funeral homes failed in 1976, or fewer than four per 10,000 (Smith and Schoen 1978c, p. 20). The average rate of failure for the industry has been calculated at one-tenth of one percent (Smith and Schoen 1978c, p. 20). Yet the most common type of funeral establishment is relatively inefficient and vulnerable to competition, whenever competition has been allowed to exist.

Traditional Establishments

The predominant type of establishment, referred to as traditional funeral establishments, is typically comprised of a wake room where services and visitation are held; a preparation room where embalming and restoration are done; a merchandise display room where caskets, vaults, clothing, and other funeral merchandise are sold; a privacy room for the family of the deceased; an office where most of the paperwork is done; and a livery where hearses and other automobiles are kept (description based on observation of 29 establishments and review of literature). The structure represents a heavy investment in fixed costs, and, coupled with the fact that funeral homes are operated on a seven-day, twenty-four hour schedule, operating costs for funeral homes are a significant portion of the cost of a funeral.

Profit margins of traditional establishments are not as lucrative as some might expect. Industry data (NFDA, various years), adjusted for inflation (1967 = 100), show that the average price of funerals more than quadrupled from 1958 to 1977, from $525 to $2,631. Pretax profit margins for roughly the same period, however, decreased from eight to only four percent of the cost of an average funeral (NFDA, various years). On the other hand, operating costs increased from 68 to 72 percent of the cost of a funeral for the same period. Other data (Federated Funeral Directors Association, various years) report that

operating costs climbed steadily from a low of 60.51 percent in 1947 to 69 percent in 1981. The latter figures are lower because they represent firms with greater volume, which tend to operate more efficiently. An additional 23 to 24 percent of the funeral price is accounted for by merchandise and remained the same over the period analyzed. In short, the average percentage of profits derived from individual funerals has fallen due to increases in operating costs.

Table 1, derived from survey data provided by the NFDA, shows that about half the establishments perform 100 or fewer funerals per year. According to U.S. Census of Business data, the average number of funerals conducted yearly by firms was 91.7 for 1967 and 96.8 for 1977 (National Foundation of Funeral Service 1982). So in terms of actual income, most funeral establishments generate profits well below the $35,000 figure implied in a recent *Wall Street Journal* (1985) report and the profit rate has been declining over time.

The revenue data also give some idea of the small size of most establishments. The Federal Trade Commission (1978, p. 43) estimates that 80 percent of all funeral firms have fewer than seven employees. Receipts of the fifty largest firms accounted for only about 8.2 percent of the total in the late 1970s, with half of this coming from the four largest firms (U.S. Bureau of the Census 1977, pp. 1–1138).

Figure 1 shows the number of funeral establishments in existence for the years 1900 through 1980. A typical pattern for popula-

Table 1. Funeral Establishments by Volume of Business: Selected Years

Year	100 or Less	100–199	200–299	300 +
1959	61%	27%	7%	5%
1960	61	27	7	5
1962	57	30	8	5
1964	56	29	10	5
1965	58	28	10	5
1966	54	31	10	5
1967	53	31	10	6
1968	49	33	12	6
1969	51	33	10	6
1973	53	27	12	8
1974	50	30	11	9
1975	52	30	11	7
1980	52	30	10	8

Note: Data from National Funeral Directors Association Annual Surveys.

Table 2. Single and Multiunit Establishments in Funeral Service

	1958	1967	1977
Total	20,767	20,191	19,622
1 establishment	18,270	18,529	16,583
2 establishments	1,764	350	1,488
3 establishments	432	189	675
4–5 establishments	221	366	373
6–10 establishments	50	494	167
11 or more establishments	30	263	336

Note: Data from U.S. Census of Businesses.

tion growth is sigmoidal (roughly an extended S-shape), with growth increasing beyond the carrying capacity of a niche, followed by a decrease to the level of the carrying capacity. The growth pattern in Figure 1 is only a partial graph, with the tail end of the "log phase" (explosive growth) in evidence during the period before 1920. A sharp drop in the 1930s coincides with the second wave of legislation, empowering funeral directors with professional autonomy. Since that time, the number of funeral establishments has fluctuated from a low of about 18,000 in 1939 to a high of approximately 22,000. The National Funeral Directors Association estimated that the number of funeral homes remained almost constant for the 10 years prior to 1985, at approximately 22,000 firms (*Wall Street Journal* 1985, p. 25). Figure 1 shows only slight variations in population size since the early 1950s.

Given the low failure rate, much of the fluctuation shown in the figure can be attributed to merger activity. Some of this activity is reflected in Table 2, comparing single and multiunit firms for 1958, 1967, and 1977. Experts in the field claim that the industry is operating at overcapacity. Other data suggest that many firms are operating at marginal levels and are kept alive because of very little intraindustry competition. A government survey conducted in the Washington, DC, area in 1973 found that comparable "least expensive" funerals ranged from $210 to $900 (Smith and Schoen 1978b, p. 16). In this same study, the average price ranged from $500 to $1830. Such pricing "undoubtedly permits many low-volume, low-capacity-utilization establishments which operate at higher average costs per service to cover these costs plus an adequate return on investment to allow them to continue in business" (p. 16). Another study (Blackwell

Fig. 1. Deaths and Funeral Homes in Existence: Selected Years
Note: Mortality data from U.S. Department of Health, Education, and Welfare, various years. Funeral home data from U.S. Bureau of the Census, various years.

1966, p. 58), focusing directly on the effects of professionally sponsored legislation on intraindustry competition found that pricing in the industry "must be such that the small firms can operate at a position on their curve compatible with stability."

Organizations Excluded From the Environment

In contrast to the stability of the traditional funeral homes and their relatively competition-free environment, three other types of funeral establishments have had tumultuous histories within the same niche. Two forms, chain establishments and cemetery/mortuary combinations, appeared early in the history of the industry, were well received by the public, but were nevertheless few in numbers as long as professional autonomy was strong. One other type, the direct-disposal firm, appeared after significant deprofessionalization in the industry.

Chain organizations were introduced in the early 1900s by an individual whose aim was to establish ". . . 100 stores, each store taking care of five funerals a week, and I am going to carry it out or die in the attempt (Necker, as cited in The Casket 1919a, p. 10)." He managed to get to the 30th "store" before going heavily into debt. The idea, however, was so attractive that, before the Depression, two of the five richest men in America

publicly expressed an interest in investing in chains (Mertens 1933, p. 17).

Despite the interest, only two prominent firms were operating in the Los Angeles area through the 1930s, 1940s, and 1950s. Pierce Brothers owned and operated six establishments by 1941 (Casket and Sunnyside 1941a, p. 74); twelve by 1950 (Mortuary Management 1950, p. 10); and twenty by 1961 (Mortuary Management 1962b, p. 8). Its rival, Utter-McKinley, had one establishment in 1936 (Mortuary Management 1950, p. 10), sixteen by 1950 (Mortuary Management 1950, p. 10), fourteen in 1961 (Mortuary Management 1962a, p. 8), and eighteen in 1964 (Mortuary Management 1964, p. 39).

It was not until the late 1960s that the funeral home chains quite suddenly appeared in greater numbers. Corporations like Uniservice, Oakridge Holdings, Kinney, American Funeral Homes, Baird-Case Funeral Homes, Hargrem Services Corporation, Olympus Service, International Funeral Services (IFS), and Service Corporation International (SCI) began to grow at phenomenal rates. The three largest in 1969, SCI, IFS, and Oakridge, had combined revenue in 1969 of $27 million, up 122 percent from 1968 revenue, with $2.1 million in profit, up 175 percent over the previous year (Forbes 1970, p. 59). After a period of merger activity among the chains, SCI and IFS became the largest firms.

International Funeral Services was estab-

lished in 1967 with three establishments and had acquired forty-seven more by 1969. By 1976, IFS owned seventy-two funeral establishments, along with five mausoleums and three cemeteries (*Casket and Sunnyside* 1976, pp. 28–29). "Between 1968 and 1972, International Funeral Services grew from a $3 million company to a $28 million company" (*Forbes* 1977, p. 144).

Service Corporation International (SCI) was founded in 1963 and had acquired five establishments by 1965 (*Wall Street Journal* 1979, p. 48). By 1979, SCI owned and operated 179 mortuaries (*Wall Street Journal 1979, p. 48). SCI* netted a 7 percent after-tax profit in 1970, "which is 3 percent higher than the mean of the top 500 corporations in the United States in the last year" (President of SCI, as cited in *Mortuary Management* 1970, p. 21). In 1982, it made its biggest acquisition, buying International Funeral Services, and increasing its holdings to 289 funeral establishments, fifty-one cemeteries, and vast amounts of real estate, which in total netted the company $180 million in revenue (*Canadian Funeral News* 1982, p. 30).

Chain organizations became efficient and profitable by first purchasing a "cluster" of funeral homes in close proximity and then adding an administrative central office (*Barrons* 1971; *Forbes* 1970). The clustering strategy enabled the firm to centralize staff and coordinate equipment usage. Coupled with advantages from buying merchandise in large volume, clustering significantly reduced operating costs and increased profit margins. Officials of one firm estimated that they could increase a mortuary's pretax profit margin to 24 percent (*Wall Street Journal* 1979, p. 48). One additional feature, which contributed greatly to the industry's reluctance to allow the firms to exist, was that they were often corporately owned. Corporations, unlike traditional funeral establishments, which were owned by licensed funeral directors, could not be controlled as easily through a professional code of ethics. As a state board chairman testifying before a public hearing explained, if the state board attempted to discipline a corporation for violating professional rules by revoking the license of the funeral director committing the act, the corporation could simply go out and hire another funeral director and continue the same practice (from testimony before the Illinois Sunset Review Committee).

Another alternative form introduced early in the history of the industry was the cemetery/mortuary combination. On December 21, 1932, the first combination, Forest Lawn, applied for a license to establish a funeral home in Los Angeles County (Burlingame 1933, p. 6). After a year of struggling to get a license for operating a funeral home, which was finally settled before the California Supreme Court, Forest Lawn succeeded in doing so (*Mortuary Management* 1933b, p. 4). By the 1950s, it had become one of the largest operators in California, with firms in Glendale, the San Fernando Valley, and a stand-alone mortuary in Long Beach (*Mortuary Management* 1962b, p. 8). While profitable, combinations were insignificant in number from the 1930s to the 1950s. As with the funeral chains, their numbers suddenly exploded in the 1960s. By 1970, there were approximately eighty-five (*Mortuary Management* 1974, p. 12). By the early 1980s there were an estimated 300 combinations (interview with NEDA official).

Like the chains, cemetery/mortuary combinations made the provision of funeral service more efficient. In this case, the need for transportation to the cemetery was virtually eliminated and, with it, the costs for vehicles and traffic control. The most significant aspect of the combination, however, was that it combined two previously independent businesses. This was problematic because the owner of the new establishment was often a cemetery owner, who did not necessarily subscribe to the professional code of ethics of funeral directors.

As funeral chains and mortuary/cemetery combinations began to appear in greater numbers, another form of structure appeared—direct-disposition firms. As Federal Trade Commission officials (1978, p. 82) described them, these firms were "designed to bypass both the mortuary and the cemetery." As the name implies, these firms offered minimal-service, minimal-cost services, usually cremation and spreading of the ashes. At one firm, bodies were described as lying on shelves, three deep, wrapped in white plastic, with the person's name written with a black marker, awaiting cremation (*American Funeral Director* 1982a, p. 8). Structurally, direct-disposition firms were a clear departure from the traditional funeral home, offering a bare minimum of services.

Telophase, the largest firm, was founded in 1971 and had conducted 3,000 cremations by 1976 (Federal Trade Commission 1978, p. 82). By that year, it had 18,000 members and six offices located on the West Coast (*Mortuary Management* 1976, p. 25). Membership in the firm involved payment of a small fee that entitled individuals to later services at low prices, usually under $500. Although accounting for only a small share of total funerals, perhaps because of its specialized niche, demand for direct disposition services is increasing (Smith and Schoen 1978a, p. 24). An accurate estimate of their population size was not available at the time of this writing.

The foregoing description of forms in the industry shows that entry for funeral home chains, cemetery/mortuary combinations, and direct-disposition firms was significantly deterred until the late 1960s. The sudden increase in nontraditional establishments occurred simultaneously with a loss of professional autonomy in the industry.

PROFESSIONALISM AND EFFECTS ON CONSTRAINING VARIATION

Exclusionary effects for three of the four types of organization described above can be traced to professionalism. Professional autonomy affected the structuring of funeral-industry firms in two ways: (1) a strategy of establishing standards that explicitly defined proper structure; and (2) a strategy of reactive exclusion. Because professionalism relies so heavily on rigid standards of practice, control of its membership is an absolute necessity. Exclusionary actions were prompted by a desire to maintain control of occupational practices.

Affecting Structure Through Standards

Blackwell found that more than 80 percent of his sample of thirty states had specific requirements for equipment and facilities to be maintained at all establishments: ". . . licensing of a funeral home may include only the payment of a fee for registration or it may go so far as to enumerate precisely the facilities that must be maintained and may provide for detailed inspection of the facilities" (Blackwell 1966, p. 174). Nearly all the states required a preparation room, with basic to detailed specifications of what further

equipment was necessary (p. 187). Other states further required a display room and an adequate stock of caskets (p. 187). Still others required a chapel, hearse, or funeral coach (p. 187) or a full-time embalmer (p. 195). Sixty-three percent of Blackwell's sample of states required licensing of funeral establishments.

An indirect method for affecting structure was licensing (Blackwell 1966, p. 181). Some states required that applicants specify the firm where they would be employed; the license would only be valid while employed at that firm (p. 181). The effectiveness of this strategy was alluded to in testimony of the Illinois Cemetery Association's legal counsel before the Illinois Sunset Review in 1982:

> . . . in order to become a funeral director, you have to be accepted as an employee of a funeral director. There was a situation several years ago . . . where a cemetery owner wanted to put up a funeral home on his property. After an extensive court case it was found that he could do so if he got a license. It became very difficult to get a license because nobody would employ him as a trainee (from taped proceedings, 1982).

Affecting Structures Through Reactive Exclusion

In a study popularized by state examinations of licensing during the 1980s, Blackwell (1966, p. 192) concluded that the unofficial "reason for the prohibition [of chain establishments] may be to prevent large funeral establishments from realizing such economies of scale that they would threaten the position of the smaller competition." Blackwell found the following laws in existence in his study: prohibition of corporate ownership in four states; requirement that all funeral establishments be fully equipped in eight states; prohibition of the sharing of facilities by establishments that did not normally maintain their own complete facilities in six states; and a requirement that all establishments employ a full-time embalmer in three states. These laws attenuated advantages of centralized equipment and staff. Coupled with licensing requirements of funeral establishments and funeral directors, they were entry barriers in many states and. effective prohibition in others. Since they were reactive in nature, often appearing after a chain organization in a given state, the threat of exclusionary laws in states without them was still present. Industry

archives are replete with descriptions of new chain organizations and effective strategies used against them. In defense of the industry, many chain organizations were so profit-motivated that they resorted to unscrupulous means for increasing revenue. For example, after successful penetration by chains in the 1960s, a former FTC lawyer and advocate of making the industry competitive contended that "if indeed there were any savings and economies in chain ownership, they weren't resulting in lower prices. It was all going to profits for [the chain]" (*Wall Street Journal* 1979, p. 48). In 1978 the Federal Trade Commission charged the same chain with overcharging customers on cash-advance items, making misleading claims about the airtightness, watertightness, and preservative capabilities of sealer caskets, requiring the purchase of caskets for cremation, embalming without permission, and offering kickbacks to hospital employees and public officials to acquire cases (Federal Trade Commission 1978, p. 37). The company's own investigation revealed that illegal payoffs to hospital and morgue employees, police, and other individuals had been made and that top executives were living high on company expense (*Wall Street Journal* 1979, p. 48). The Georgia Department of Consumer Affairs charged another chain with using bait-and-switch tactics, involving the offering of free plots in an attempt to lure customers into buying more expensive ones (ATCOM 1982, p. 11).

In response to Forest Lawn in 1933, the California state board quickly sought to introduce legislation prohibiting the form (*Mortuary Management* 1933a, pp. 5–6). At a National Funeral Directors Association regional meeting in San Francisco, articles were drafted for inclusion in the Code of Ethics (Berg 1933, p. 2) prohibiting the operation of cemetery/mortuary combinations, prohibiting the sale of funeral service in conjunction with the sale of burial plots, and prohibiting the use of salesmen employed by a cemetery for selling funeral services. In Illinois, lawyers were hired to establish a legal precedent that would prohibit combinations based on "the theory that tax-exempt cemeteries should not be permitted to compete with tax-paying funeral directors and also on the theory that a cemetery cannot be licensed to operate a mortuary (*Mortuary Management* 1962b, p. 38)." A similar strategy was attempted in

Pennsylvania (*Mortuary Management* 1963, p. 46). Other states used laws that prohibited solicitation because, "these firms are not just a 'cemetery/mortuary;' but rather a cemetery and a mortuary plus door-to-door selling" (based on interview with association official). A Federal Trade Commission investigation (1978, p. 119) found that twelve states explicitly prohibited the establishment of combinations, while others relied on more indirect strategies of exclusion.

Direct-disposition firms appeared at a time when professionalism for the industry was declining, making it more difficult for the industry to influence the legislative process. The Federal Trade Commission reported that some states did attempt to prohibit the forms, while others sought to bring them under the authority of state boards, where they would be subjected to already-existing laws that could be used to exclude them. They also found other evidence of efforts to eliminate the threat of direct-disposition firms: boycotting crematories that did business with Telophase, one of the most visible of the direct-disposition firms; boycotting ambulance-service firms that did business with Telophase; refusal by some companies to do business with Telophase because it would hurt their chances of doing business with traditional establishments; efforts to pressure the county health director to refuse requests for death certificates from non-funeral directors; and efforts to prevent the granting of authority for Telophase to operate its own crematory. Thus, when professional power was waning, collective action was brought in as an alternative force. The unity and homogeneity of funeral directors as a profession were instrumental in mobilizing for protective action against outsiders even when their legal autonomy was decreasing.

LOSS OF PROFESSIONALISM LEADING TO SUCCESSFUL PENETRATION BY ALTERNATIVE FORMS

The foregoing evidence supports the thesis that professional autonomy enhanced the industry's ability to constrain variation. This final section presents more evidence of the relationship between professionalism and variation, obtained from sources outside of the industry.

Beginning in the middle 1960s, consumer groups targeted the funeral industry for needed reform, spurred by best sellers like

Jessica Mitford's *American Way of Death* (1963), Ruth Harmer's *High Cost of Dying* (1963), and a host of investigations by journalists (see, for example, *Business Week* 1963; *Business Week* 1964; *Changing Times* 1963; *Consumer Reports* 1964; *America* 1976). Beginning in the 1970s, the federal government and some states responded by conducting their own investigations.

The Federal Trade Commission began a preliminary inquiry on funeral-industry practices in December of 1972, citing evidence of wrongdoing such as the following (FTC 1978): performing services not requested by the client; requiring that customers purchase items that were not legally necessary; misrepresenting legal, public health, or religious requirements for the purpose of increasing sales of merchandise; failure to provide adequate price lists of the various services and merchandise available and/or disclosing these prices by telephone; practices hindering the sale of lower-priced funerals; policies prohibiting the advertising of prices; and failure to keep accurate records of business transactions. The FTC approved a rule for consideration in 1979; however, passage of the FTC Improvement Act of 1980, giving Congress the power to veto any proposed FTC rule, further limited the scope of FTC rule-making (*Chicago Sun-Times* 1982, p. 37). The FTC then decided to delay submission of the rule to Congress until such time as these limitations could be complied with. Another version of the rule was written and approved on July 22, 1981 (Federal Trade Commission 1982), a much watered-down version of the original due to strong opposition by the industry. The rule was submitted to Congress in June 1982 (National Selected Morticians 1982) and ultimately approved in 1984. The FTC rule, in general, required that funeral directors make explicit the cost of various items and services, provide pricing information over the telephone, and refrain from implying that certain services and items were legal requirements (*Washington Post* 1984, p. 30). A subtle but significant change had taken place — official recognition of the need for increased competition in the industry and legitimation of structures that would accomplish this.

Individual states also began a series of "sunset reviews," examinations of the need for professional licensing, in the 1970s. Of fourteen states conducting such investigations, Colorado, New Hampshire, Texas, Connecticut, and Kansas recommended abol-

ishing professional status for the industry (Illinois Select Joint Committee on Regulatory Agency Reform 1982, p. 27). Of these, Colorado was the only state to do so (*American Funeral Director* 1982b, p. 32). During the course of investigation, however, the states discovered that the industry had used its professional autonomy "to upgrade the funeral directing profession, not to combat consumer abuse" (Illinois Bureau of the budget 1982, pp. 26–27; Kansas Legislative Division of Post Audit 1980, p. 30).

While actual loss of professional autonomy occurred in only one state, the result of so much political attention effectively decreased professional autonomy. New laws proposed by industry leaders had little chance of passage and existing laws were questioned. Compelled by a new awareness that much existing regulation sought merely to protect traditional establishments rather than consumers, new state regulation aimed at making the field more competitive. Occurring simultaneously with the loss of autonomy was successful penetration by chains, combinations, and direct-disposition firms. In summary, important discontinuous changes of a political nature allowed variation to exist where it had previously been excluded.

Individuals within the industry suggested that consumer advocacy was only a smokescreen and that the real forces behind deprofessionalizing were interests connected with alternative structures. "Until the direct disposers started to operate in California about ten years ago, there were few who questioned funeral service license laws and licensing boards," contended officials from the National Funeral Directors Association (1981, p. 30). Others pointed out that the large firms themselves were behind a substantial proportion of the legal activity that tested the constitutionality of professional legislation (interviews with NFDA and other trade-association officials). Regardless of where the change originated, the tie between professionalism and constraints on building and operating alternative forms is clear.

DISCUSSION AND SUMMARY

This paper has argued that professionalism was used to influence variation in both a proactive and reactive sense; that is, the profession defined acceptable structures and eliminated potential variation within its envi-

ronment. It is clear that the strategy resulted in increased survival chances for one type of organization, but the interesting dynamics occur at the level of constraining variation, not at the level of selection.

While one case study cannot generalize to all professions, and even less to all institutions, other work does support the observations made here. Most major works on professionalism support the general notion that the more developed a profession, the greater the control it has over its environment. Moreover, the relationship discussed here addresses one of the major criticisms of population ecology—that it does not work well with cases where special interests have some control over their environments.

A central question is whether population-ecology theory is flexible enough to deal with institutional constraints. Hawley's criticism, that ecology theorists have tended to over-study the selection process, is in line with the response offered here. Hawley's criticism is not of the theory itself, but, rather, with the way it is used. I argue similarly that to deal with cases of imperfect competition, population ecology must first do away with the assumption that variation is any less important than selection. Because variation precedes selection, the key to understanding imperfect competition may lie in a deeper understanding of what affects this process. In essence, population ecology should go beyond the question, "What contributes to an increase or decrease of net mortality for given populations?" to include the question, "What contributes to the presence or absence of variation?"

In many ways, the analytical approaches that were used to introduce the ecological perspective to the social sciences forced the field to gravitate toward the study of selection. Natural ecology, from which organizational ecology derives its theoretical framework, preceded analytical work with many years of observation, even if today this activity "is pooh-poohed by hotshot young theorists who regard it as the dog work of unimaginative drones" (Gould 1977, p. 128). Indeed, Darwin's major works, *The Origin of Species* (1859) and *The Voyage of the Beagle* (1845), are heavily observational and oriented toward the study of variation. Students of natural ecology were much better equipped to deal with the question of variation than are organizational ecologists. This is not to say

that observational research should replace analytical research, but rather that population ecology should pay closer attention to the questions "Who is competing?" and "Why?" In some cases, this may require closer observation of the variety of organizations competing, those wanting to compete, and the origins of each.

There are several theoretical frameworks that can be brought to bear on the question, "What contributes to the presence or absence of variation?" In the area of professionalization, for example, "process theories" (see, for example, Johnson 1972; Larson 1977; Starr 1982; Friedson 1970) deal almost exclusively with the issues of controlling environments and implications for the survival of given types of organizations. More generically, the theory of "structuration" (Giddens 1976; 1982) operates from a general assumption that social actors have an "effective though incomplete knowledge of themselves and a limited capacity to 'make a difference' to the course of events" (Dickie-Clark 1986, p. 168). These theories can complement, not replace, the ecological perspective.

In summary, population ecology has proven to be a useful tool for understanding survival in relatively competitive environments; however, it has been criticized as unable to deal with the case of imperfect competition. As part of the solution, this paper calls for extending its focus to the variation process, noting that, in some environments, variation is not random. In cases where it is not, the most provocative aspects of population ecology are at the level of variation, not selection.

REFERENCES

Aldrich, Howard. 1979. *Organizations and Environments*. Englewood Cliffs, NJ: Prentice-Hall.
America. 1976. "Cost of Being Buried: Justice Department Antitrust Suit." 117 (December): 703–4.
American Funeral Director. 1982a. "Telophase Backstage." 105(April):8.
_____. 1982b. "Editorial Opinion: Comes Sunset, Out Goes the Baby." 105(June):32.
Astley, W. Graham. 1985. "The Two Ecologies: Population and Community Perspectives on Organizational Evolution." *Administrative Science Quarterly* 30:224–41.
Astley, W. Graham and Andrew Van de Ven. 1983. "Central Perspectives and Debates in Organizational Theory." *Administrative Science Quarterly* 28:243–73.

ATCOM. 1982. Funeral Service Professional's Handbook. New York: ATCOM Publishing.

Barrons. 1971. "Undertaker to Entrepreneur." (August 2):11,15,17.

Berg, William. 1933. "Editorial: NFDA Must Use Code to Keep Cemeteries Out." Mortuary Management 20(August):2.

Betton, John and Gregory G. Dess. 1987. "The Application of Population Ecology Models to the Study of Organizations." Academy of Management Review 10,4:750–57.

Blackwell, Roger D. 1966. "Price Levels of Funerals: An Analysis of the Effects of Entry Regulations in a Differentiated Oligopoly." Ph.D. diss., Northwestern University.

Brown, Richard E. 1979. Rockefeller Medicine Men. Medicine and Capitalism in America. Berkeley: University of California Press.

Burlingame, George E. 1933. "Operation of Mortuaries by Interment Organizations Viewed as a Menace." Mortuary Management 20(February):6,8.

Business Week. 1963. "Burial Business Hears Voices: Concerning Jessica Mitford's American Way of Death." (November):80–84.

———. 1964. "Checking the Burial Bills: Senate Hearings on Burials." (July 11):36.

Canadian Funeral News. 1982. "Funeral Industry Brightens." 91(January):30–31.

Carroll, Glenn R. 1984. "Organizational Ecology." Annual Review of Sociology 10:71–93.

Carroll, Glenn R. and Yangchung Paul Huo. 1986. "Organizational Task and Institutional Environments in Ecological Perspective: Findings from the Local Newspaper Industry." American Journal of Sociology 91:838–73.

Casket, The. 1919a. "Pike's Peak or Bust." XLIV(May):10.

Casket and Sunnyside. 1941a. "Public Responsive to Service Innovations." 71(April):74.

———. 1941b. "Portland, Oregon, Embalmers Union in Dispute with Teamsters Union." 71(December):50.

———. 1948. "Embalmers' Strike Averted in Manhattan." 78(January):70–71.

———. 1976. "IFS buys Bunker Home and Cemeteries." 106(September):28–29.

Changing Times. 1963. "Furor over Funerals: Why the Cry of Scandals?" T17(November):7–12.

Chicago Sun-Times. 1982. "Itemized Listing of Funeral Prices Voted by FTC." (July 29):37.

Colorado Department of Regulatory Agencies. 1980. Report of the State Auditor. Denver, CO: State of Colorado, Office of State Auditor.

Consumer Reports. 1964. "Concerning Three Books on the Funeral Industry." 29(January):40–43.

Crouch, Ben M. 1971. Professionalism in Funeral Service: A Study of Work Orientations. Ph.D. diss., Southern Illinois University.

Darwin, Charles. 1845. The Voyage of the Beagle. Geneva, Switzerland: Heron Books.

Darwin, Charles. 1859. The Origin of Species. New York: Literary Classics, Inc.

Delacroix, Jacques and Glenn R. Carroll. 1983. "Organizational Foundings: An Ecological Study of the Newspaper Industries of Argentina and Ireland." Administrative Science Quarterly 28:274–91.

Dickie-Clark, H.F. 1986. "The Making of a Social Theory: Anthony Giddens' Theory of Structuration." Sociological Focus 19:159–76.

DiMaggio, Paul J. and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." American Sociological Review 48:147–160.

Federated Funeral Directors of America. Various Years. Annual Surveys. Reports of the Management Analysis Division, Federated Funeral Directors of America, Springfield, IL.

Federal Trade Commission. 1978. Funeral Industry Practices: Final Staff Report to the FTC and Proposed Trade Regulation Rule. Washington, DC: Department of Consumer Protection.

———. 1982. FTC Factsheet on Proposed Funeral Industry Practice Rule. Washington, DC: Department of Consumer Protection.

Forbes. 1970. "Merchants of Death." 106(November):59.

———. 1977. "Consumer Resistance." 113(November):144.

Freeman, John and Michael T. Hannan. 1983. "Niche Width and the Dynamics of Organizational Populations." American Journal of Sociology 88:1116–45.

Freeman, John, Glenn R. Carroll, and Michael T. Hannan. 1983. "The Liability of Newness: Age Dependence in Organizational Death Rates." American Sociological Review 48:692–710.

Friedman, Milton. 1962. Capitalism and Freedom. Chicago: University of Chicago Press.

Freidson, Eliot. 1970. Professional Dominance: The Social Structure of Medical Care. New York: Atherton Press.

Giddens, Anthony. 1976. New Rules of Sociological Method. New York: Basic Books.

———. 1982. Profiles and Critiques in Social Theory. London: Macmillan.

Gilligan, Harry. J. 1928. "The Value of Organization." Mortuary Mangement 15 (May):16–17.

Gould, Stephen J. 1977. Ever Since Darwin: Reflections in Natural History. New York: W.W. Norton.

Habenstein, Robert and William M. Lamers. 1955. The History of American Funeral Directing. Milwaukee, WI: Bulfin Printers

———. 1984. "Structural Inertia and Organizational Change." American Sociological Review 49:149–64.

———. 1987. "Organizational Founding and

American Labor Unions." *American Journal of Sociology* 92:910–43.

Harmer, Ruth M. 1963. *The High Cost of Dying.* New York: Collier, MacMillan.

Hawley, Amos. 1981. "Human Ecology." *American Behavioral Scientist* 24:423–44.

Herringer, J. 1976. *The Funeral Service Industry: An Economic, Marketing and Financial Investigation.* Merrick, NY: Morton Research Company.

Hirsch, Paul M. 1986. "From Ambushes to Golden Parachutes: Corporate Takeovers as an Instance of Cultural Framing and Institutional Integration." *American Journal of Sociology* 91:800–37.

Hughes, Everett C. 1965. "Professions." Pp. 1–14 in *The Professions in America,* edited by Kenneth Lynn. Boston: Houghton Mifflin.

Illinois Bureau of the Budget. 1982. Governor's Sunset Report on the Illinois Funeral Directors and Embalmers Act. Springfield, IL: Illinois Bureau of the Budget.

Illinois Select Joint Committee on Regulatory Agency Reform. 1982. Briefing Materials: An Act in Relation to the Regulation of Persons Engaged in the Practices of Funeral Directing and Embalming. Unpublished report. Springfield, IL: Illinois Select Joint Committee on Regulatory Agency Reform.

Johnson, Terence J. 1972. *Professions and Power.* London: MacMillan Press.

Kansas Legislative Division of Post Audit. 1980. Sunset Audit Report: Board of Embalming. Report submitted to the Legislative Post Audit Committee, State of Kansas, September, 1980.

Larson, Magali Surfatti. 1977. *The Rise of Professionalism: A Sociological Analysis.* Berkeley, CA: University of California Press.

Lubove, Roy. 1980. *The Professional Altruist: The Emergence of Social Work As a Career, 1880–1930.* New York: Atheneum.

Lynn, Kenneth S. 1965. *The Professions in America.* Boston: Houghton Mifflin Company.

McKelvey, Bill and Howard Aldrich. 1983. "Populations, Natural Selection, and Applied Organizational Science." *Administrative Science Quarterly* 28:101–28.

Mertens, Fred C. 1933. "Wall Street is Looking Your Way." *Mortuary Management* 20 (January):17–25.

Meyer, John W. 1978. "Strategies for Further Research: Varieties of Environmental Variation." Pp. 352–68 in *Environments and Organizations,* edited by Marshal Meyer and Associates. San Francisco: Jossey-Bass.

Meyer, John W. and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83:340–63.

Mitford, Jessica. 1963. *The American Way of Death.* New York: Simon and Schuster.

Mortuary Management. 1928. "The Burial Survey Completed." 15 (February):20–22.

———. 1929a. "Purge the Profession of the Unfit." 16 (March):33.

———. 1929b. "Announcing National Plans." 16 (June):22.

———. 1933a. "Funeral Director-Cemetery Compromise Improbable." 20 (March):5–6.

———. 1933b. "Forest Lawn Conducting Funerals." 20 (December):4.

———. 1950. "Death is Big Business in Los Angeles County." 37 (July):10,12,14,42.

1951. "Editorial: The Forty Hour Week." 38 (September):12.

———. 1952. "Los Angeles Union Gets Upset." 39 (April):30.

———. 1953. "San Francisco Embalmers' Strike Threatened." 40 (April):34.

———. 1956. "Mortuaries, Cemeteries Strikes Continue to Plague San Francisco." 43 (July):6.

———. 1962a. "Editorial: Chains Eliminate You? No!" 49 (April):8.

———. 1962b. "Illinois Fights Cemetery/Mortuary." 49 (May):38.

———. 1963. "Minnesota Experiences Legislative Activity." 50 (April):46.

———. 1964. "Utter-McKinley Adds 18th." 51 (September):39.

———. 1970. "Managing Editor William M. Berg Talks with Robert L. Waltrip." 57 (December):17–23.

———. 1972. "Recognition for the American Board of Funeral Services." 59 (April):42.

———. 1974. "Death Among the Dying." 61 (March):12–13.

———. 1976. "Interview with Thomas B. Weber." 63 (September):25–29.

———. 1978. "An Interview with Arthur R. Angel." 65 (April):29–31, 37–38.

National Foundation of Funeral Service. 1982. "Basic Funeral Service Data." Unpublished report, Evanston, IL: NFFS.

National Funeral Directors Association. Various Years. Proceedings of the Annual Convention. Milwaukee, WI: National Funeral Directors Association.

———. Various Years. A Statistical Abstract of Funeral Service Facts and Figures of the United States. Milwaukee, WI: National Funeral Directors Association.

———. 1981. *Tradition in Transition: The Report of the 21st Century Committee of the National Funeral Directors Association.* Milwaukee, WI: NFDA.

National Selected Morticians. 1982. "FTC Update." NSM Newsletter, June 10.

Neilson, William and C. G. Watkins. 1973. Proposals for Legislative Reforms Aiding the Consumer of Funeral Industry Products and Services. Burnsville, NC: Celo Press.

Pfeffer, Jeffrey. 1978. "The Micropolitics of

Organizations." Pp. 29–50 in *Environments and Organizations*, edited by Marshal Meyer and Associates. San Francisco: Jossey-Bass.

———. 1982. "The External Control of Organizational Behavior." Pp. 178–204 in *Organizations and Organization Theory*. Marshfield, MA: Pitman.

Perrow, Charles. 1986. *Complex Organizations*. New York: Random House.

Renouard, Charles A. 1944. "Professional Competence Impossible Without Science." *Casket and Sunnyside* 74 (February):32–36.

Singh, Jitendra, David J. Tucker, and Robert J. House. 1986. "Organizational Legitimacy and the Liability of Newness." *Administrative Science Quarterly* 31:171–93.

Smith, Ronald G.E. and Gerald L. Schoen. 1978a. "Trends in Funeral Service." *Mortuary Management* 65 (January):23–25,46.

———. 1978b. "Competition and Regulation." *Mortuary Management* 65 (February):15–18,37.

———. 1978c. "Some Implications of Chain Operation and Investor Ownership of Funeral Homes." *Mortuary Management* 65 (May):20–23.

Starbuck, William H. 1983. "Organizations as Action Generators." *American Sociological Review* 48:91–102.

Starr, Paul. 1982. *The Social Transformation of American Medicine*. New York: Basic Books.

Stinchcombe, Arthur L. 1959. "Bureaucratic and Craft Administration Production: A Comparative Study." *Administrative Science Quarterly* 4:169–87.

———. 1965. "Organizations and Social Structure." Pp. 153–93 in *Handbook of Organiza-*

tions, edited by James G. March. Chicago:Rand-McNally.

Texas Sunset Advisory Commission. 1978. Staff Report to the Sunset Advisory Commission. Austin, TX: State Board of Morticians.

Torres, David L. 1983. "An Industry and its Environment: Professionalism and its Effects on the Funeral Industry." Ph.D. diss., Northwestern University.

Tushman, Michael L. and Philip Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31:439–65.

U.S. Bureau of the Census. Various years. *Census of Service Businesses*. Washington, DC: U.S. Department of Commerce.

———. 1985. *Statistical Abstract of the United States*. Washington, DC: U.S. Department of Commerce.

U.S. Department of Health, Education, and Welfare. Various years. *Vital Statistics of the United States:* vol. II—Mortality, Part B. Hyattsville, MD: National Center for Health Statistics.

Wall Street Journal. 1979. "SCI Reaches Top of the Funeral Industry But Its Practices Have Raised Questions." (December 6):48.

———. 1985. "The Most Likely to Survive—A Funeral Home or a Florist?" (September 16):25.

Washington Post, National Weekly Edition. 1984. "The Truth-In-Funerals Requirements." 1 (May 14):30.

Wiewel, Wim and Albert Hunter. 1985. "The Interorganizational Network as a Resource: A Comparative Case Study on Organizational Genesis. *Administrative Science Quarterly* 30:482–96.

SHAME AND CONFORMITY: THE DEFERENCE-EMOTION SYSTEM*

THOMAS J. SCHEFF

University of California, Santa Barbara

This article proposes a unitary explanation of social control for normal and rigid conformity. Conformity may arise from the interaction of deference with normal pride and shame; rigid conformity from chain reactions of shame. I show that Darwin, Cooley, and others suggested the same context for pride and shame: self's perception of the evaluation of self by other(s). Their work, which assumes a continuous social monitoring of the self from the standpoint of others, suggests a puzzle: If social monitoring is continuous and causes either pride or shame, why are so few manifestations of either emotion visible in our lives? One possible explanation is that pride and shame usually have very low visibility. I call this the Cooley-Scheff conjecture. Goffman's work on "face" implies this conjecture and Lewis's discovery of unacknowledged shame confirms it. Her analysis of hundreds of clinical interviews demonstrates that low-visibility shame was present in every session, though neither therapist nor patient seemed to be aware of it. Drawing on Lewis's exact description of the markers of various manifestations of shame and Goffman's analysis of the relation between deference and embarrassment, a deference-emotion system is described. Members perceive this system as compelling conformity to norms exterior to self by informal but pervasive rewards (outer deference and its reciprocal, inner pride) and punishments (lack of deference, and the inner shame that is its reciprocal). I show how Asch's study of conformity and independence illustrates the role of shame in compelling conformity to exterior norms.

Durkheim ([1897] 1951) bequeathed to modern social thought a theoretical building block: the idea that the force of social influence is experienced by individuals as *exterior and constraining*. Although he argued that the individual experiences social influence as an absolutely compelling force exterior to self, he did not spell out the causal sequence implied. What are the steps that lead individuals to experience social control as exterior and constraining? This is an important question because exterior constraint has become a basic premise for modern sociologists. Yet, an adequate model has never been conceptualized, much less proposed in an operationally definable way. Conformity poses

a central problem for social science not only in its normal, but also in its pathological, form. What gives rise to excessive and rigid conformity? This is the question posed by the many modern analyses of bureaucracy and authoritarian forms of social organization. In this article I will outline a model that speaks to both forms of the problem.

There is wide agreement that conformity is encouraged by a system of sanctions: we usually conform because we expect to be rewarded when we do and punished when we do not. However, conformity usually occurs even in the absence of obvious sanctions. Durkheim's formulation refers to the ubiquity of conformity. The reward of public acclaim and the punishment of public disgrace rarely occur, yet the social system marches on. Formal sanctions are slow, unwieldy, and expensive. In addition to the formal system, there must be a complex and highly efficient system of informal sanctions that encourages conformity.

A clue to this puzzle can be found in Goffman's treatment of interaction ritual (Goffman 1967). He notes that *the emotion of*

*embarrassment or anticipation of embarrass-
ment plays a prominent role in every social
encounter.* In presenting ourselves to others,
we risk rejection. The form rejection takes
may be flagrant, but it is much more
frequently quite subtle, perhaps only a missed
beat in the rhythm of conversation. Depend-
ing on its intensity and obviousness, rejection
usually leads inevitably to the painful emo-
tions of embarrassment, shame, or humilia-
tion. By the same token, when we are
accepted as we present ourselves, we usually
feel rewarded by the pleasant emotions of
pride and fellow feeling.

I propose that the degree and type of
deference and the attendant *emotions* of pride
and shame make up a subtle and pervasive
system of social sanctions. This system leads
to experiencing social influence as compel-
ling. Our thoughts and perceptions of social
expectations only set the stage for social
control. We experience the system as so
compelling because of emotions—the plea-
sure of pride and fellow feeling on the one
hand, and the punishment of embarrassment,
shame, or humiliation on the other.

The deference-emotion system may take
formal and public forms: the ceremony for
awarding the Congressional Medal of Honor
confers the highest degree of deference and
may be assumed to arouse pride in the
recipient. At the other extreme, an impeach-
ment proceeding takes away deference and
presumably would arouse shame in the
defendant. Disgrace subsumes both public
and private sides—outer demotion and inner
shame.

But formal rewards and punishments are
infrequent, even rare. The deference-emotion
system functions virtually continuously, even
when we are alone, since we can imagine and
anticipate its motions in vivid detail. System-
atic research has been unable to document this
system; it is too subtle and ubiquitous for
laboratory experiment or social survey. Since
it often functions outside of awareness,
qualitative fieldworkers have not yet been
able to catch the details.

Unlike the system of formal sanctions, the
deference-emotion system is virtually instan-
taneous and invisible. Its invisibility makes it
difficult to describe. Although shame and
pride are implied in Durkheim's writing about
social influence, they are never named.
Shame appears to be profoundly taboo
(Scheff 1984); it is not mentioned even when

it is being used as an explanation. Asch's
discussion of his findings (see below) illus-
trates this evasiveness in a modern study of
conformity.

Even Goffman, whose analysis of the role
of embarrassment is a tour de force, falls far
short of a complete analysis. Although he
points to the ubiquity of embarrassment in
social encounters, he restricts his purview to
the *social* aspects of embarrassment, to what
is going on *between* interactants, and ex-
cludes any psychological components (1967,
p. 108).

Limiting his analysis to outer behavior in
the social domain pays rich dividends in
certain areas, e.g., the contagion of embar-
rassment between interactants. Goffman's
treatment of social embarrassment is subtle
and evocative; but since he does not attempt
to explain the interaction between outer and
inner processes, he is unable to convey the
explosive force of the deference-emotion
system. One difficulty is that his analysis
completely separates embarrassment ("the
social organization of embarrassment") from
anger and hostility, which he treats in
"character contests" such as duels and
vendettas ("where the action is") (1967).

It is constructive to contrast Lewis's (1971)
treatment of shame with Goffman's, since it
is the photographic negative of his. She treats
only the inner process. In analogy to his use
of the metaphor of contagion between per-
sons, she points to what she calls a *feeling
trap,* i.e., inner contagion. In Goffman's
analysis, one becomes ashamed that the other
is ashamed, who in turn becomes ashamed,
which increases the first person's shame, and
so on—an *interpersonal* feeling trap. In
Lewis's analysis, one becomes ashamed that
one is ashamed, an inner loop which feeds on
itself—an *intrapersonal* trap. Unlike Goff-
man, however, she does not separate her
analysis of shame from her analysis of anger.
She postulates an affinity between the two
emotions, with shame usually followed by
anger. This loop, which can go on indefi-
nitely, is usually experienced as though it
were a single affect, "helpless anger," or, in
a more intense form, "humiliated fury."

By combining Goffman's social analysis
with Lewis's psychological one, it is possible
to convey the extraordinary power of the
deference-emotion system. This system oc-
curs both *between* and *within* interactants.
Ordinarily it functions so efficiently and

invisibly that it guarantees the alignment of the thoughts, feelings, and actions of individuals. Mutual conformity and respect lead to pride and fellow feeling, which lead to further conformity, which leads to further positive feeling, in a system that seems virtually automatic.

However, when there is a real and/or imagined rejection on one or both sides (withdrawal, criticism, insult, defeat, etc.) the deference-emotion system may show a malign form, a *chain reaction* of shame and anger between and within the interactants. This explosion is usually brief, perhaps a few seconds. But it can also take the form of bitter hatred and can last a lifetime. It can occur not only between individuals, but also between groups, or even nations. I refer to such explosions as *triple spirals* of shame and anger (one spiral within each party and one between them). A chain reaction between and within groups can last longer than a lifetime, handed down from generation to generation; I interpret Franco-German relations (1870–1945) as an extended spiral of this kind (Scheff 1987; forthcoming).

For all its brilliance, Goffman's analysis of interaction ritual implies that such matters may be fateful, at most, only to individuals, but not in larger arenas. Embarrassment, he seems to imply, can be exquisitely painful, but it is personal and transitory and not relevant to larger social institutions. His behavioral analysis, which excludes the psychological domain and separates embarrassment from anger, is too specialized to capture the larger implications of his vision.

Lewis's specialization, equal but opposite to Goffman's, also precludes her from drawing out the social implications of her work. Although she is aware that her concept of the feeling trap has implications beyond neurosis, there is little development in this direction in her written work. Only by combining her and Goffman's partial analyses can their implications be seen.

Because of the ubiquity of shame and shame/anger sequences, *all* social and societal interaction can instantly become what Goffman calls a character contest. When chain reactions of shame or shame/anger occur between and within interacting persons or groups, there is no natural limit to the intensity and duration of arousal. *The unlimited fury of shame/rage in a triple spiral may explain why social influence can be experi-*

enced as absolutely compelling. The emotion-deference system, as represented in the sequence of honor, insult, and revenge, may decide the fate, not only of individuals, but of nations, civilizations, and, in our nuclear era, of all life on earth.

By analysis of *sequences* of interaction ritual and emotion in concrete episodes, it may be possible to enlarge on Durkheim's [1897] 1951) investigation of suicide. In another place, I outline a model of the way in which a class-based insult led to suicide in a classic work of fiction (Scheff and Mahlendorf 1988). As mentioned above, I also interpret the bizarre and highly self-destructive behavior of France and Germany in terms of the interaction ritual between the two countries. French and German politics and diplomacy from 1870–1945 appear to be extremely irrational and perhaps may be understood as a character contest, even at the risk of self-destruction. With analyses of the interaction between deference and emotion, it may be possible to develop Durkheimian ideas of social influence into a comprehensive study of interaction at both the interpersonal and institutional levels.

THE SOURCES OF SHAME: BIOLOGICAL AND SOCIAL

In modern societies shame is considered rare among adults. This belief is reflected in the division made in anthropology between shame cultures and guilt cultures, with traditional societies relying on shame for social control, and modern societies, guilt. A matching premise is found in orthodox psychoanalytic theory, which places almost total emphasis on guilt as the adult emotion of self-control, with shame thought of as "regressive," that is, childish. (An early attempt to break away from both restrictive premises can be found in Piers and Singer [1953].)

For many years, however, there has been a continuing suggestion in the literature that shame is the primary social emotion, generated by the virtually constant monitoring of the self in relation to others. Such monitoring, as suggested in my reference to Goffman, is not rare but almost continuous in social interaction, and, more covertly, in solitary thought. *If this line of thought is correct, shame would be the most frequent and possibly the most important of emotions, even*

though it is usually almost invisible. Threads of this idea can be found in Darwin (1872), Cooley (1922), MacDougall (1908), and, more recently, in Lynd (1958), Lewis (1971), and Goffman (1967).

In *The Expression of Emotions in Men and Animals* (1872), Darwin devotes a whole chapter to blushing and its relation to shame. He states his thesis quite simply: blushing is caused by "shyness, shame, and modesty, the essential element in all being self-attention." For my purposes here, the important proposition comes next in his text, where he explains what he means by self-attention: "It is not the simple act of reflecting on our own appearance, but the *thinking what others think of us, which excites a blush*." (emphasis added, p. 325) His discussion suggests that blushing may be caused by perceptions of other people's evaluation of the self, whether positive or negative.

Darwin's argument about the relationship between blushing and self-attention can be restated as two propositions connecting blushing with what might be called, in current terms, emotions, on the one hand, and social perception, on the other. First, *blushing is caused by shame* (as discussed below, "shyness" and "modesty," Darwin's two other causes of blushing, can be considered shame variants [Lewis 1971] or cognates [Wurmser 1981]). Second, and more importantly, *shame is caused by the perception of negative evaluations of the self*. Blushing is only one of several visible markers of overt shame, and therefore is not a primary concept for a theory of social influence. The second statement, however, contains the basic proposition for the whole theory: shame is *the* social emotion, arising as it does out of the monitoring of one's own actions by viewing one's self from the standpoint of others.

Shame as a crucial emotion for adults is prominent in the work of William MacDougall (1908). He thought of shame as one of the "self-regarding sentiments," perhaps the most important one: "Shame is the emotion second to none in the extent of its influence upon social behavior" (p. 124). Like Darwin, he seems to have understood that it arises as a result of self-monitoring. He also makes another important point, that, although shame undoubtedly has a biological basis that we share with the higher mammals, the human emotion of shame in adults is considerably more elaborate and complex (p. 56).

THE COOLEY-SCHEFF CONJECTURE

We next turn to Cooley (1922), who considers pride and shame as the crucial "social self-feelings." At some points he seems to regard as a self-feeling *any* feeling that the self directs toward itself. This passage about the extraordinary importance of self-feelings in human behavior seem to be in this key: ". . . with all normal and human people, (social self-feeling) remains, in one form or another, the *mainspring of endeavor and a chief interest of the imagination throughout life*." (emphasis added, p. 208)

Cooley continues:

> As is the case with other feelings, we do not think much of it (that is, of social self-feeling) so long as it is moderately and regularly gratified. Many people of balanced mind and congenial activity scarcely know that they care what others think of them, and will deny, perhaps with indignation, that such care is an important factor in what they are and do. But this is illusion. If failure or disgrace arrives, if one suddenly finds that the faces of men show coldness or contempt instead of the kindliness and deference that he is used to, he will perceive from the shock, the fear, the sense of being outcast and helpless, that he was living in the minds of others without knowing it, just as we daily walk the solid ground without thinking how it bears us up. (p. 208)

Although neither pride nor shame is mentioned in this passage, they are implied, *especially the almost continuous presence of low-visibility pride in ordinary discourse*. Cooley thought of pride and shame as the crucial self-feelings.

This idea is continued in the concept of "the looking-glass self," his description of the social nature of the self. He thought self-monitoring has three steps:

> As we see our face, figure, and dress in the glass, and are interested in them because they are ours, and pleased or otherwise with them according as they do or do not answer to what we should like them to be; so in imagination we perceive in another's mind some thought of our appearance, manners, aims, deeds, character, friends, and so on, and are variously affected by it.
>
> A self-idea of this sort seems to have three principal elements: the imagination of our appearance to the other person; the imagination of his judgment of that appearance, and some sort of self-feeling, such as pride or mortification. (p. 184)

In this passage he restricts self-feelings to the two he seems to think are the most significant, pride and shame (considering mortification to be a shame variant). He mentions shame three more times in the passage that follows (emphasis added):

> The comparison with a looking-glass hardly suggests the second element, the imagined judgment, which is quite essential. The thing that moves us to *pride or shame* is not the mere mechanical reflection of ourselves, but an imputed sentiment, the imagined effect of this reflection upon another's mind. This is evident from the fact that the character and weight of that other, in whose mind we see ourselves, makes all the difference with our feeling. We are *ashamed* to seem evasive in the presence of a straightforward man, cowardly in the presence of a brave one, gross in the eyes of a refined one, and so on. We always imagine, and in imagining share, the judgments of the other mind. A man will boast to one person of an action—say some sharp transaction in trade—which he would be *ashamed* to own to another. (pp. 184–85)

What is unfamiliar about the looking-glass self, perhaps shockingly so, is that Cooley is implying that society rests on a foundation of pride and shame. His analysis of the social nature of the self can be summarized in two propositions:

1. In adults, social monitoring of self is virtually continuous, even in solitude. (We are, as he put it, "living in the minds of others without knowing it.") (208).
2. Social monitoring always has an evaluative component, and gives rise, therefore, to either pride or shame.

Together these propositions suggest a puzzle. If social monitoring of self is almost continuous, and if it gives rise to pride or shame, why do we see so few manifestations of either emotion in adult life? Among possible answers is that the pride or shame is there, but has such low visibility that we do not notice it. This answer gives rise to a third proposition, which I will call the Cooley-Scheff conjecture:

3. Adults are virtually always in a state of either pride or shame, usually of a quite unostentatious kind.

This proposition is a step toward an exact definition of a concept that has been so far undefined: *level of self-esteem.* Such a definition would concern the balance between pride and shame states in a person's life, taking into account both duration and intensity. This issue will be the topic of a future article.

In his discussion of grief (he calls it distress-anguish), Tomkins (1963) notes a parallel puzzle: "The reader must be puzzled at our earlier affirmation that distress is suffered daily by all human beings. Nothing seems less common than to see an adult cry. And yet we are persuaded that the cry, and the awareness of the cry, as distress and suffering, is ubiquitous." (p. 56)

His answer also parallels the one I have suggested: "The adult has learned to cry as an adult. It is a brief cry, or a part of a cry or a miniature cry, or a substitute cry, or an active defense against the cry, that we see in place of the infant's cry for help." (p. 56) He goes on to discuss various substitutes for, or defenses against, crying that adults employ. For example, an adult suffering in the dental chair might, instead of crying, substitute muscular contractions: clamping the jaw, tightly contracting the muscles in the abdomen, and rigidly gripping the arms of the chair (p. 59). As an example of defending against the cry, Tomkins suggests masking the facial expression of sadness with one of anger, becoming angry as well as sad (pp. 64–65). Most men in our society use this transformation, but many women do the opposite, masking anger with grief.

One way of summarizing the gambits that adults use when they are suffering is that most adults' grief is of a type with low visibility because its manifestations have been disguised or ignored. Tomkins' question and answer with respect to adult grief are exactly parallel to the ones I have derived from Cooley's treatment of adult shame.

What may be the most dramatic of Cooley's views on shame, and the one which brings him closest to my position is his use of an autobiographical excerpt to illustrate the power of what he calls "social fear":

> Social fear, of a sort perhaps somewhat morbid, is vividly depicted by Rousseau in the passage of his *Confessions* where he describes the feeling that led him falsely to accuse a maid-servant of a theft which he had himself committed. "When she appeared my heart was agonized, but the presence of so many people was more powerful than my compunction. I did not fear punishment, but I dreaded *shame:* I dreaded it more than death, more than the crime, more than all

the world. I would have buried, hid myself in the center of the earth: invincible *shame* bore down every other sentiment; *shame* alone caused all my impudence, and in proportion as I became criminal the fear of discovery rendered me intrepid. I felt no dread but that of being detected, of being publicly and to my face declared a thief, liar, and calumniator. (p. 291, emphasis added)

Rousseau's phrase, "invincible shame," will stand us in good stead in the reinterpretation of the Asch study I undertake below. Notice also that Cooley suggests this instance is an example of "morbid" (i.e., pathological), rather than normal, shame. I use a similar distinction in my discussion of the Asch experiment.

Cooley's discussion of the social self in terms of self-monitoring (the movement now called "role-taking") clearly invokes pride and shame as the basic social emotions. At this point, intellectual history takes a surprising turn. G.H. Mead and John Dewey based virtually their entire social psychology on the process of role-taking, the ability of humans to continuously monitor their own selves from the point of view of others. Yet neither Mead nor Dewey ever mention what was so obvious to Darwin, MacDougall, and Cooley—social monitoring gives rise to feelings of pride or shame. Mead and Dewey treat role-taking, their basic building block of human behavior, as almost entirely a cognitive process. Neither has anything to say about pride and shame, as if Darwin, MacDougall, and Cooley never existed. Social psychology has yet to recover from this oversight.[1]

In modern societies, adults seem to be uncomfortable manifesting either pride or shame. *The emotions of shame and pride often seem to themselves arouse shame.* (This proposition explains Darwin's observation that both positive and negative evaluations can give rise to blushing.) It seems likely, as both Darwin and MacDougall suggest, that shame has a biological basis and is genetically programmed, not just in humans, but in the higher mammals. It may also be true, as recent infant-caretaker studies suggest, that for infants and very young children, the

arousal of shame is largely biological. For adults, however, it also seems certain that shame is not only a biological process, but also an overwhelmingly social and cultural phenomenon. The discussion so far has suggested that adult shame is doubly social: shame arises in social monitoring of the self, and shame itself often becomes a further source of shame, depending on the particular situation and normative structure of the culture.

The second social aspect of shame, its recursiveness, can give rise to *pathological* shame, a potentially limitless spiral, (Scheff 1987). As will be suggested below, the concept of pathological shame may explain the Asch effect, and more broadly, all excessive or rigid conformity.

LOW-VISIBILITY SHAME

If, as I have suggested, shame is strongly recursive in modern societies, we would expect most shame and pride to have very low visibility. Even if shame and pride were widely prevalent, persons who were proud or ashamed would be ashamed of their state and attempt to hide it from others and from themselves. If this is the case, how can one study pride and shame if they are usually hidden?

I know of no systematic studies of pride, but a beginning method for detecting low-visibility shame was developed by Gottschalk and Gleser (1969). Their procedure for extracting emotions from verbal texts includes a long list of sentences containing words they consider shame markers. These sentences are listed under five categories (I provide a few of their examples under each category.):

1. Shame, embarrassment: "I feel funny . . ." ". . . I had behaved improperly. . ." (and other sentences using terms such as *disconcerting, discredit,* or *unworthy*).
2. Humiliation: "I don't know what was wrong with me letting myself go like that . . ." (other sentences involve such terms as *humbling, degrading,* or *little self-respect*).
3. Ridicule: "He twitted me about being fat . . ." ". . . I really feel utterly ridiculous in a situation like that . . ." ". . . They stared at me and laughed . . ."
4. Inadequacy: "Where was I when brains were passed out? I feel stupid . . ."
5. Overexposure of deficiencies or private details: "I don't even know how to wipe my ass.

---

[1] A prior attempt to rectify Mead and Dewey's oversight can be found in Shibutani (1961), particularly chap. 13, "Self-esteem and Social Control," which implies, in part, the thesis of the present article.

I didn't want to talk about such personal things." (pp. 49–52)

Although Gottschalk and Gleser do not discuss the matter or refer to any of the shame theorists discussed above, a few of the sentences contain explicit references to shame. Instead, most of their examples assume what the shame theorists posited to be the basic context for shame—a perception of negative evaluation of the self by self or others, even if the negative evaluation is somewhat indirect. Nor do the authors attempt to include nonverbal markers of shame.

In her pioneering analysis of clinical dialogues, Lewis (1971) takes up the issue of shame markers much more explicitly and broadly than Gottschalk and Gleser. Her work is both theoretical and empirical, tying broad concepts and hypotheses to concrete episodes of behavior. In this capacity she is the heir of Darwin, MacDougall, and Cooley. She advances our knowledge of shame, however, because, unlike either the original theorists or the more recent advocates (Lynd [1958], Tomkins [1963], and Goffman [1967]) who use carefully selected examples only in an illustrative way, she conducted a systematic analysis of shame content in complete episodes of real social interaction: entire clinical sessions. Her laborious, word-for-word analysis of these sessions led her to the discovery of what she calls "unacknowledged" shame, the low-visibility shame predicated here.

Lewis first distinguishes acknowledged and unacknowledged shame. She shows that in hundreds of clinical sessions, most of the shame episodes were virtually invisible to the participants, unacknowledged by either the patient or the therapist. She divides these episodes into two basic types: *overt, undifferentiated* shame and *bypassed* shame.

Overt, undifferentiated shame involves painful feelings that are not identified as shame by the person experiencing them. These feelings are instead referred to by a wide variety of terms that disguise the shame experience: feeling foolish, stupid, ridiculous, inadequate, defective, incompetent, awkward, exposed, vulnerable, insecure, having low self-esteem and so on.

Lewis classified all these terms as shame markers because each occurred in conjunction with (1) *contexts* in which the patient appeared to perceive self as negatively evaluated, either by self or other(s), the

central context for shame; and (2) a change in the patient's *manner,* characterized by nonverbal markers such as speech disruption (stammering, repetition of words, speech "static" like "well," "uhhhh," long pauses, etc.), lowered or averted gaze, blushing, and, especially noticeable, a sharp drop in volume, often resulting in inaudibility.

Both the verbal and nonverbal markers of overt shame can be characterized as forms of *hiding* behavior. The verbal terms disguise shame, and the nonverbal forms suggest physical hiding: averting or lowering the gaze to escape the gaze of the other, hiding behind a mask-like blush, and hiding the meaning of speech and thoughts behind speech disruption and inaudible speech.

To summarize, overt, undifferentiated shame occurs when a person (1) feels the self negatively evaluated, either by self or other; (2) manifests hiding behavior (speech disruption, lowered or averted gaze, blushing, or barely audible speech); and (3) labels or associates the painful feeling with undifferentiated terms such as those listed above. In these instances, although the negative evaluation of self appears so painful as to interfere with the fluent production of thought and/or speech, the pain is mislabeled.

In addition to the overt, undifferentiated pattern, Lewis describes the second pattern of unacknowledged shame, *bypassed* shame. Like the overt pattern, bypassed shame always begins with a perception of the negative evaluation of self. Where the markers of undifferentiated shame are flagrant and overt, those of bypassed shame are subtle and covert. Although thought and speech are not obviously disrupted, they take on a speeded-up, repetitive quality that Lewis refers to as *obsessive.*

Typically, patients repeated a story or series of stories, talking rapidly and fluently, but not to the point. They appeared to be unable to make decisions because of seemingly balanced pros and cons ("insoluble dilemmas"). Patients complained of endless internal replaying of a scene in which they felt criticized or in error. Often they report that when they first realized the error, they winced or groaned, then immediately became obsessed with the incident. The mind seems to be so taken up with the unresolved scene that one is unable to become adequately involved in the present, even though there is

no obvious disruption. One is subtly distracted.

The two patterns of shame appear to involve opposite responses. In overt, undifferentiated shame, victims *feel* emotional pain to the point that it obviously retards or disrupts thought and speech. They seem to be trying to hide the painful state from themselves as well as from others. In bypassed shame, the victim appears to *avoid* the pain through hyperactive thought, speech, or actions. These two types appear to correspond to my own distinction between under- and over-distanced emotion (Scheff 1979). Overt, undifferentiated shame is under-distanced, since the intense pain of embarrassment or humiliation is experienced. What Mead (1934) called the "I" phase of the self, the "biologic individual," predominates in consciousness. Bypassed shame is over-distanced; one avoids the pain by stepping outside of self, into the "me" phase of the self, as if the pain were not happening.

Adler's (1956) theory of human development anticipated Lewis's discovery of the two basic types of unacknowledged shame. Although he did not use the term, what he called "the feeling of inferiority," i.e., shame, played a central role in his theory. He argued that developing children's greatest desire is for love. If love is not available at the crucial points, the development of their personality can proceed along one of two paths. Either they develop "inferiority complexes," i.e., they become prone to overt, undifferentiated shame, or they compensate by seeking power, i.e., they avoid feeling shame by bypassing it, through what I have termed hyperactive thought, speech, or actions.

Both the slowed-down pattern of overt shame and the speeded-up pattern of bypassed shame are disruptive, however, because both involve the victim in rigid and distorted reactions to reality. Both kinds of shame are equally invisible; one is misnamed, the other ignored. These two basic patterns explain how shame might be ubiquitous, yet usually unnoticed.

The work of Tomkins and Lewis converge with and extend Freud's early work on repression. In his first book (1897, with Breuer), he argued that hysteria was caused by repressed emotion, "strangulated affect," as he called it. He observed that patients improved when they expressed forgotten emotions, e.g., by crying or using heated words, a rudimentary theory of catharsis (Scheff 1979).

Tomkins (1963) approached repression from a very different direction, through deduction about the fate of grief that did not result in catharsis (crying). This approach led him to describe the outer signs of low-visibility grief, an affect that overlaps with, but is not identical with, that of repressed grief (some low-visibility grief is a result of conscious or partly conscious masking or avoidance).

Lewis extended the concept of repression both theoretically and empirically. Using the kind of shuttling back and forth between deductive and inductive methods that Peirce called *abduction* (Scheff, forthcoming), she laid the groundwork for the shame construct, the description of the context and markers for unacknowledged shame, and its role in the genesis and maintenance of neurotic symptoms. Most of the shame episodes she reported seem to be not only unperceived but not available to consciousness. Her work therefore confirms not only the Cooley-Scheff conjecture (the part about shame), but also confirms and expands on Freud's ([1897] 1966) original hypothesis that neurosis is caused by strangulated affect.[2] I will now turn to the reanalysis of a study that points to the role of shame in rigid or excessive conformity, a much more widely prevalent form of pathological behavior than traditional psychopathology.

THE ASCH CONFORMITY STUDIES

Asch's (1956) study of conformity illustrates the way in which emotions may lead to social control. In this classic laboratory study, single subjects found themselves alone, facing what seemed to be a unified majority. Since the task was a simple comparison of the lengths of lines, the naive subjects must have been baffled by the completely erroneous responses of the other subjects. Unknown to the naive subject, the others were confederates of the investigator, instructed to give erroneous responses. A high proportion of conforming behavior resulted: three-quarters of the subjects were swayed at least once by the majority responses, only one-quarter remained completely independent.

In this study, Asch followed the inductive

---

[2] In a parallel finding, Volkan (see Volkan and Josephthal 1979 for a list of citations) discovered the syndrome of pathological grief, and its cure, a cathartic treatment, called "regrief therapy."

design that characterizes modern experimental social psychology; it is not a test of an hypothesis derived from theory. In retrospect, however, the study comes close to testing a basic aspect of Durkheimian theory by holding constant one element in social influence (exteriority) while allowing the other element (constraint) to vary. All of the naive subjects perceived the judgment of the majority to be different from, and therefore exterior to, their own judgment. Asch's study demonstrates the exteriority of group standards by showing that naive subjects who were allowed to make judgments alone, without an erring majority, made no errors at all.

The study tests the hypothesis that *given the kind of task demanded, a majority of the subjects will find group standards compelling, even though they are exterior and contradictory to their own individual standards.* This formulation suggests a key question: what is the difference between those subjects who maintained their independence from the group and those who yielded? I will suggest that the answer involves the part played by emotions: subjects who yielded to the majority were attempting to avoid the embarrassment (shame) of appearing different from the group.

Although Asch did not design the study to show the effects of emotion or ask questions about emotions, many of the post-study interview responses suggest that emotions played an important part in the results. It is clear from both Asch's observations and from direct analysis of the subject's responses that many of them found the experience of being in the minority extremely painful. Ash reports that as the division between the majority and the individual continued, the individual became more tense, flustered, nervous, and anxious. A reaction that occurred both in independent and yielding subjects was the fear that they were suffering from a defect that the study would disclose this defect: "I felt like a silly fool. . . . A question of being a misfit. . . . they'd think I was queer. It made me seem weak-eyed or weak-headed, like a black sheep" (p. 31).

Many of the comments show negative viewing of self from the point of view of the others: "You have the idea that the attention of the group is focused on you. I didn't want to seem different. I didn't want to seem an imbecile. They might feel I was just trying to

be out of the ordinary. . . . They probably think I'm crazy or something" (p. 31). These comments are all markers of overt shame (see the earlier discussion of markers).

Asch makes an important point, however, in differentiating the post-study interview responses of those who remained independent from those who yielded to the majority. He notes, first, that both independent and yielding subjects were troubled by disagreeing with the majority:

> As the disagreement persisted many began to wonder whether it signified a defect in themselves. They found it painful to be (as they imagined) the focus of attention, in addition to which they feared exposure of their weakness which they suspected the group would disapprove. (p. 32)

Asch points out that though these feelings were all but universal, not everyone who experienced them yielded. Those who maintained their independence responded to their own perceptions of the lines despite their strong feelings. Asch characterizes the responses of those who yielded in a different way:

> They were dominated by their exclusion from the group which they took to be a reflection on themselves. Essentially they were unable to face a conflict which threatened, in some undefined way, to expose a deficiency in themselves. They were consequently trying to merge in the group in order not to feel peculiar. (p. 45)

There are several key ideas in this passage. First, there are two markers of overt, undifferentiated shame: "feel peculiar," and "expose a deficiency in themselves." A second point is implied by the first sentence, the subjects' perception of "exclusion from the group," which they took to be a "reflection on themselves" rather than on the group. Because the group took no action to actually exclude the naive subjects, the perception of exclusion must have been solely in the subjects' imaginations, implying the basic shame context: perceiving one's self negatively evaluated.

Finally, the entire passage can be summarized by translating it into the language of emotions. The subjects who yielded to the group were those who not only felt overt shame, but whose perceptions, cognitions, and/or actions were controlled by the attempt to avoid it. Conversely, many of the subjects who maintained independence from the group

appeared to feel conflict, i.e., overt shame, but elected to hold their ground in spite of this feeling.

So far the discussion of the role of emotion in compelling conformity has been entirely in terms of overt, undifferentiated shame. There is also scattered evidence in the subjects' post-study remarks, and in Asch's summaries of these remarks, of bypassed shame as a causal element in compelling conformity. In the summary by Asch cited above, he refers to those subjects who acknowledged conflict between themselves and the group. There is another group of subjects, however, all of whom yielded to the influence of the group, in which there is little or no acknowledgement of conflict and no markers of overt shame. The emotion markers in this group suggest the presence of bypassed shame.

After stating that the interview responses of the independent subjects were apt to be frank and forthright, he notes that those of the yielding subjects were different:

[Their] reactions were more often evasive and shallow, and some revealed a lack of appreciation of the situation and of the possible significance of their action. . . . When asked to describe his experiences at the outset of the interview, one subject (who yielded completely) inquired: "Exactly what do you mean by experiences?" Another remarked: "I didn't have any experiences, I felt normal." (p. 33)

These two responses, especially the latter, suggest a complete denial of conflict and the feelings resulting from that conflict. The comments of two other yielding subjects also suggest denial, but in a different form; they reduced their experience of the study to ponderous or obfuscating generalizations: "People perceive things differently . . ." and "How do we know who is right?" Another group of yielding subjects, according to Asch,

"granted to the majority the power to see correctly . . . allowed themselves to become confused, and at the critical point adopted the majority judgments *without permitting themselves to know of their activity.*" (emphasis added, p. 42)

This is a very strong statement about avoidance of conflict and denial of feelings through self-deception. Asch found among the yielding subjects another frequent form of self-deception: considerable underestimation of the number of times each had yielded. He

showed that the amount of underestimation was proportional to the amount of yielding (the greater the yielding, the greater the underestimation) (pp. 34–35). Many of these subjects acted as if there were only a few such incidents, even those who had yielded at every opportunity. By denying reality, perhaps they sought to avoid painful feelings.

One of Asch's summaries of the interview responses emphasizes both types of shame responses—the pain of overt shame and what I have called here the obsessive quality of bypassed shame:

Our observations suggest that independence requires the capacity to accept the fact of opposition without a lowered sense of personal worth. The independent person has to organize his overt actions on the basis of experience for which he finds no support; this he can do only if he respects his experiences and is capable of claiming respect for them. The compliant person cannot face this ordeal because he translates social opposition into a reflection of his personal worth. Because he does so the social conflict plunges him into pervasive and incapacitating doubt. (p. 51)

Most of this passage is commensurate with both forms of shame, but the last sentence seems to focus on the obsessive form, the insoluble dilemma characteristic of bypassed shame. On the whole, Asch's report inadvertently suggests that unacknowledged shame plays a central role in causing subjects to yield to group influence, even when it contradicts their own direct perceptions of reality.

To conclude this section, I will return to a still unresolved issue in my interpretation. Recall that many of the subjects who remained independent, as well as some of those who yielded, manifested markers of overt, undifferentiated shame. (All of those cited above who showed bypassed shame had yielded.) In this case, the presence of bypassed shame may adequately explain yielding; however, overt, differentiated shame does not because it seemed to be present in both those who yielded and those who did not. For this latter group, at least, some further explanation is needed.

One possible formulation is in terms of self-esteem. The subjects who remained independent, although they experienced shame, had sufficiently high self-esteem to act on their judgments despite their feelings of shame. Those who yielded had low self-

esteem and sought to avoid further feelings of shame by acting contrary to their own judgment. Asch comes close to stating it in these very terms. Such a formulation, a causal explanation based on the personalities of the subjects, could be tested in future studies.

Although this formulation could mark an advance in understanding, the advance would probably be very slight. The concept of low self-esteem can be seen as a gloss, implying a person who habitually feels shame rather than pride. Perhaps the concept of pathological shame, as already indicated, would specify the causal process more precisely. Low self-esteem might be conceptualized as a tendency toward endlessly recursive shame, spirals of potentially limitless intensity and duration. As suggested in my earlier article (Scheff 1987), in such persons, shame alone, or in combination with other emotions such as anger, might be recursive to the point of chain reaction. Such a dynamic sequence could explain explosive episodes of acute panic, (a shame-fear alternation), resentment (shame-anger alternation, with the anger directed out), and guilt (shame-anger sequences with the anger directed in).

Using this model, persons with high self-esteem would be those with the experience of managing most shame so that it is acknowledged and discharged. Although shame is a painful emotion for them, as for everyone else, it is not overwhelming. Persons with low-self esteem would be those who have been unable to manage shame. For such persons, a situation that threatened to be shaming would be perceived as overwhelmingly painful, since it might involve them in an unending spiral of shame. Such a person might do anything to avoid the pain, to "turn the world upside down, rather than turn themselves inside out," to use one of Helen Lewis's favorite phrases. One example is the episode in which Rousseau's conscience lost the battle with what he called "invincible shame." Another is a troublesome and otherwise baffling remark that one of the yielding subjects made to Asch in the post-study interview. He said that he voted for Dewey in the election of 1948, even though he preferred Truman, because he thought Dewey would win and that he was preferred by most (p. 48). Apparently, unacknowledged shame is not only invincible, but insidious.

This formulation does not imply that social control is dependent entirely on individual personality, only that individuals have differing susceptibilities to shame. The situational component in conformity is equally important in the deference-emotion model (the deference awarded to individuals in a situation). The Asch experiment was almost diabolical with respect to shame because it was arranged to rob, in a covert way, the individual of her own view of reality. As the deference-emotion model suggests, conformity results from the interaction of individual and situational components. This interaction is also a cultural phenomenon because status arrangements consist, at the microscopic level of analysis, of the blend of awarding and withholding deference (for an earlier formulation, which suggests a link between status and emotion, see Kemper 1977).

SUMMARY

I have proposed that Durkheim's analysis of social influence implies a deference-emotion system in which conformity to exterior norms is rewarded by deference and feelings of pride, and nonconformity is punished by lack of deference and feelings of shame. In this analysis, social control involves a biosocial system that functions silently, continuously, and virtually invisibly, occurring *within* and *between* members of a society. Cultural taboos on the acknowledgement of pride and shame seem to lead to pathological states of shame, which give rise to rigid or excessive conformity. I have shown how Asch's study of independence and conformity may be reinterpreted in these terms. If the deference-emotion system is universal, the theory would provide a unitary explanation of conforming behavior, the central problem of social science.

REFERENCES

Adler, Alfred. 1956. *The Individual Psychology of Alfred Adler*. New York: Basic Books.

Asch, Solomon. 1956. "Studies of Independence and Conformity: 1. "A Minority of One against a Unanimous Majority." *Psychological Monographs* 70:1–70.

Cooley, Charles H. 1922. *Human Nature and the Social Order*. New York: Scribners.

Darwin, Charles. 1872. *The Expression of Emotion in Men and Animals*. London: John Murray.

Durkheim, Émile. [1897] 1951. *Suicide*. New York: Free Press.

Freud, Sigmund and Josef Breuer. [1897] 1966. *Studies of Hysteria*. New York: Avon.

Goffman, Erving. 1967. *Interaction Ritual*. New York: Anchor.

Gottschalk, Lewis A. and Goldine C. Gleser. 1969. *Manual of Instruction for Using the Gottschalk-Gleser Content Analysis Scales*. Berkeley: U. of California Press.

Kemper, Theodore D. 1977. *A Social Interactional Theory of Emotions*. New York: Wiley.

Lewis, Helen B. 1971. *Shame and Guilt in Neurosis*. New York: International Universities Press.

Lynd, Helen M. 1958. *Shame and the Search for Identity*. New York: Harcourt Brace.

MacDougall, William. 1908. *An Introduction to Social Psychology*. London: Metheun.

Mead, G.H. 1934. Mind, Self and Society. Chicago: U. of Chicago Press.

Piers, Gerhart and Milton B. Singer. 1953. *Shame and Guilt: A Psychoanalytic and Cultural Study*. New York: Norton.

Scheff, Thomas J. 1979. *Catharsis and Healing in Healing, Ritual, and Drama*. Berkeley: U. of California Press.

———. 1984. "The Taboo on Coarse Emotions."

Review of Personality and Social Psychology 5: 146–69.

———. 1987. "The Shame/Rage Spiral: Case Study of an Interminable Quarrel." Pp. 109–50 in *The Role of Shame in Symptom Formation*, edited by H.B. Lewis. Hillsdale, NJ: Lawrence Erlbaum.

———. Forthcoming. *Micro-sociology: Implicature and Emotion in Discourse*.

Scheff, Thomas J. and Ursula Mahlendorf. 1988. "Emotion and False Consciousness: Analysis of an Incident from *Werther*." *Theory, Culture, and Society* 5:57–79.

Shibutani, Tamotsu. 1961. *Society and Personality*. Englewood Cliffs, NJ: Prentice-Hall.

Tomkins, Silvan S. 1963. *Affect/Imagery/Consciousness*. New York: Springer.

Volkan, Vladimir and Daniel Josephthal. 1979. "Brief Psychotherapy in Pathological Grief: Regrief Therapy." Pp. 102–52 in *Specialized Techniques in Psychotherapy*, edited by T.B. Karasu and L. Bellak. New York: Brenner/Mazel.

Wurmser, Leon. 1981. *The Masks of Shame*. Baltimore: Johns Hopkins University Press.

MACROECONOMIC AND SOCIAL-CONTROL POLICY INFLUENCES ON CRIME RATE CHANGES, 1948–1985*

JOEL A. DEVINE
JOSEPH F. SHELEY
M. DWAYNE SMITH
Tulane University

Informed by the insights of political economy, this study investigates the often-presumed though empirically elusive relationship between societal economic distress and crime. In a social-indicators model, we argue for including both unemployment and inflation rates as measures of the overall health of the economy. We contend that in the face of these destabilizing economic conditions, government engages in dualistic social control policies. On one hand, it attempts to discourage antisocial behavior via placative forms of control and, on the other, it exercises its deterrent capacities. Using annual time-series data for the period 1948–1985, we employ dynamic modeling techniques to examine these influences on annual fluctuations in rates of homicide, robbery, and burglary. The results yield mixed support for the hypothesized relationships, with the posited model gaining potency as we move from explaining more violent to less violent offenses. Finally, these findings hold when we control for changes in two other theoretically important influences: criminal opportunity and the age structure of the population.

Land and Felson (1976) have argued that sociologists should develop models of the complex social processes that generate reported crime rates. More recently, Cohen and Land (1987) contend that relatively simple models can account for a substantial proportion of the variance in annual rates of certain types of crime. In this paper, we build such a model through time-series analyses of changes in systemic economic distress, governmental attempts to offset that distress and deter criminal behavior occasioned by it, and annual fluctuations in reported rates for homicide, robbery, and burglary for the years 1948–1985. Our model incorporates both generative and inhibitory elements. However, we extend traditional attention to these elements by arguing for the inclusion of both

inflation and unemployment rates as indicators of the overall health of the economy. We contend as well that models that contain a social-control element, as ours does, must address both deterrent (e.g., criminal justice) and placative (e.g., welfare) forms of control.

ECONOMIC DISTRESS AND CRIME

Though the search for links between specific societal economic conditions and crime rates has an extensive history in the sociological literature, the relationship remains ambiguous (Chiricos 1987; Long and Witte 1981). However, no matter the theoretical orientation, at the aggregate level economic distress is viewed as a precipitator of negative social conditions that undermine legitimacy and order and weaken social bonds. While theorists disagree about how economic distress influences individuals to commit particular forms of crime (Thornberry and Christenson 1984), they generally agree that systemic economic distress increases the probability of higher crime rates for the system in question. Regarding unemployment rates, Cantor and Land (1985) state:

> Array the members of a population along [a] continuum according to their levels of [criminal] motivation. . . . [A]ll other things being equal, an increase in the unemployment rate produces a

shift in the density distribution of the population along this continuum towards its higher end. Thus, the central tendency (mean, median) of the motivation density will have shifted upward. Furthermore, if it is assumed that the level of crime experienced by the population is an (unconditional) increasing function of the level of the central tendency of this density distribution, then it follows that the crime rate should increase. (p. 319)

Similarly, we contend that inflation raises a population's criminal motivational density. In a situation in which many wage and salary earners and persons on fixed or minimum wage incomes either experience an erosion of real income or simply perceive such a loss (see Peretz 1983, chap. 3-4), inflation unleashes distributional conflict and undermines confidence in existing institutional arrangements (Hirsch and Goldthorpe 1978; Lindberg and Maier 1985).[1] Moreover, while inflation fosters anomie and fuels a general climate of uncertainty and fear (see Foster 1981), it simultaneously limits government's (fiscal) capacity to maintain both deterrent

---

[1] Methodological issues aside, economists still are not agreed on the distributional effects of inflation. On one hand, researchers have discovered a pro-poor (i.e., progressive) distributional impact (see, e.g., Hollister and Palmer 1972; Minarik 1979; Devine and Canak 1986). On the other hand, evidence indicates that, at least in the post-1967 period, the price of basic necessities rose faster than the cost of luxuries, suggesting that the less affluent suffered greater economic harm (Peretz 1983, pp. 54-55; Mirer 1973). Arguably, any true inflationary effects were at least partially mitigated for social security recipients and organized workers by cost-of-living adjustments (COLAs). However, public assistance payments and the minimum wage were not protected and their recipients experienced a reduction in real income. Though the overall long-term distributive effects of inflation are relatively minor, the bottom 20 percent of income recipients (as well as the top 5 percent) have experienced some income loss in the period under study. In the short term, inflation appears to have had a generally negative impact, especially on wage and salary incomes (Peretz 1983, chap. 3). Whatever their differences concerning the "true" effects of inflation, however, few social scientists would disagree with Peretz's (1983, p. 234) claim that most people "tend to overstate or misperceive the negative effects of inflation." Whether or not accurate, the perception of inflation as having negative effects is what motivates behavior and, in turn, structures our thesis.

and placative forms of control (O'Connor 1973, especially pp. 243-45).

Models of the relationship between societal economic conditions and crime must use both inflation and unemployment rates as indicators of economic distress (or health). Both unemployment and inflation critically shape macroeconomic and social-welfare policies. Throughout the postwar period, governmental policy makers have recognized the potential economic and social consequences of unemployment and inflation and consciously have sought to manage both rates to maintain overall economic health and foster sociopolitical stability.

SOCIAL CONTROL

A host of governmental policies (both anticipatory and reactive) are directed at the troubled economy and the problems emanating from it. We view governmental attempts at inhibiting criminal activity in general terms: governments counter crime with a carrot-and-stick approach in which they seek to placate "problem populations" thought to possess high rates of deviant behavior and to prevent, through threat and punishment, members of that population from engaging in such behavior. Governments indirectly address the espoused link between economic distress and crime through the provision of welfare while simultaneously and directly pursuing deterrent and incapacitative criminal-justice policies.

In our view, research regarding economic conditions, social control, and crime insufficiently attends to this dual nature of control. The majority of studies of crime-rate change over time address the social-control issue only in terms of deterrent and incapacitative criminal-justice policy (see, for example, Cantor and Land 1985). However, the claim that public relief serves as an instrument to control troublesome classes characterizes a considerable literature concerning class and social problems in the U.S. (Colvin and Pauly 1983; Piven and Cloward 1971).

Conservative critics of crime-control efforts in the United States view relief as ineffective in reducing crime and argue that the failure to control the "criminal classes" severs already fragile bonds to the conventional work ethic (Wilson 1985, pp. 117-44). In response, liberals claim that a wide range of social programs for the poor may reduce delinquency

rates among those persons left behind by conventional economic growth (Currie and Skolnick 1984, pp. 447–48). In this vein, it is noteworthy that the late-1960s produced both the President's Commission on Law Enforcement and Administration of Justice, which argued that crime must be prevented by programmatic efforts to relieve poverty and inequality, and the Omnibus Crime Control and Safe Streets Act, which sought to put teeth into law enforcement in the face of public panic about society's "crime problem." In sum, models aimed at explaining crime-rate changes since World War II must take into account placative control.

MODEL AND HYPOTHESES

We postulate a model of macroeconomic and social-control policy influences on crime-rate changes that uses unemployment and inflation rates as indicators of economic conditions that may be viewed as more or less distressful over time (see Figure 1). Social-control policy is indicated by the prison-population rate and by the amount of public relief provided to the poor. The former was chosen as best reflective of the "lock 'em up" deterrence and incapacitation anticrime policies of the 1970s and 1980s; the latter represents what many observers view as the traditional policy response to social problems arising in market-based economies.

Like virtually all aggregate-level theories about the link between economic distress and/or social-control policies and the rate of crime (see Long and Witte [1981] for an overview of this literature), our conceptualization of this process entails a theory of change. Transformations in the economy

(e.g., unemployment) or public-policy provisions (e.g., deterrent activities) are thought to provoke changes in the crime rate. Consequently, we employ dynamic modeling techniques (time-series difference equations) to estimate the effects of *changes* in the independent variables on *changes* in the rates of homicide, robbery, and burglary.

As Figure 1 indicates, we posit positive relationships between the three index-crime rates and the two indicators of macroeconomic fluctuation and negative relationships between crime rates and social-control policies.

INDEPENDENT VARIABLES

Unemployment. Despite the traditionally intuitive assumption of a positive link between aggregate unemployment and crime rates, the most recent sophisticated research into the matter indicates a more complicated relationship. Cantor and Land (1985) report both positive and negative effects of unemployment on crime: a negative partial effect across five types of violent and property offenses (homicide, robbery, burglary, larceny, and motor vehicle theft) and a significant positive partial effect for robbery, burglary, and larceny alone. They argue that the criminal motivational aspects of unemployment produce the positive effects, while the criminal opportunity (victim-target availability) aspects account for the negative findings. For the latter three crimes mentioned above, the total impact of unemployment-rate changes on crime-rate changes is negative. More recently, Cohen and Land (1987) report a negative effect of unemployment on motor vehicle theft rates and a positive effect on homicide rates. The differences between their
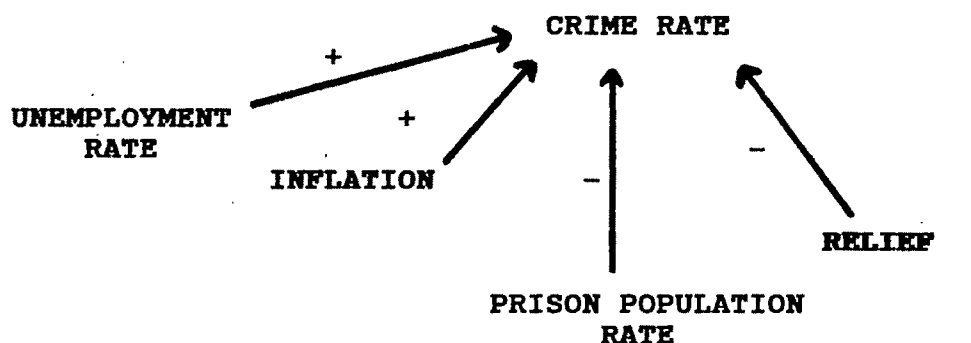
Fig. 1. Postulated Model of Macroeconomic and Social Policy Influences on Changes in the Rate of Crime

findings and those of Cantor and Land (1985) are attributable to modeling and time-series length inconsistencies between the two studies. Further underscoring the complexity of the issue at hand, Parker and Horwitz (1986) examine the unemployment/crime relation using yet another analytical strategy. They report little evidence of a link between the two variables at the state level for the years 1974–1979.[2]

Given the similarity of our approach and data to those of Cantor and Land (1985), their study informs our hypothesis; however, we conceptualize the relation of unemployment to crime in terms of unemployment's motivational component. Hence, we posit a positive relation between the two variables in our model. We later address the opportunity aspect of unemployment by entering a measure of opportunity into the equations examined below.

Inflation. Generalizing from the literature on inflation and white-collar crime and on income inequality and violent and property offenses, Long and Witte (1981, p. 134) posit an increase in crime rates as inflation rises— both because hard times motivate criminal behavior and because inflation inhibits the capacity of communities to deter crime. Beyond this, Brenner (1976) reports a positive relationship between homicide rates and inflation from 1940–1973 (but see Cohen and Felson 1979a), and Land and Felson (1976) report a statistically significant positive relationship between inflation and property (but not violent) crime rates for the years 1947–1972. These findings are clearly in line with the general literature's theme that economic distress begets criminal activity and are sufficient to inform our hypothesis that inflation-rate changes are positively related to crime-rate changes.

Prison Population Rate. The standard state response to criminal behavior is the imposition of criminal penalties on offenders. A considerable literature addresses the effects of the rate of imprisonment on rates of criminal activity, focusing on the inhibitory (i.e., deterrent) effect of the degree of the perceived risk of imprisonment (Zedlewski 1983) and/or the incapacitative effects of imprisoning serious offenders (Cohen 1983). While research on these themes is not definitive, the thesis underlying them is sufficiently clear to structure our hypothesis: increased imprisonment rates will have a negative effect on crime rates.[3]

Relief. As noted, little attention has been devoted to the relation of placative social-control policy to crime rates. Conservative claims notwithstanding, the overwhelming theme in the literature characterizes public assistance and other welfare programs as aimed at quieting those who might react antagonistically (presumably criminally) to their adverse economic status (see, e.g., Currie 1985, pp. 104–41). In one of the few empirical studies of the effect of relief on crime, DeFronzo (1983) reports that the level of public assistance provided the poor was inversely related to crime rates in 39 sampled SMSAs. We therefore posit a negative relationship between changes in amounts of relief to the poor and crime-rate changes.

Control Variables. Two themes in the recent criminological literature suggest the need to introduce two control variables into the test of our model. The first variable is age structure of the population. An extensive literature suggests a link between crime rates and the proportion of the population in high crime-prone age categories (Cohen and Land 1987; Steffensmeier and Harer 1987). Therefore, we later enter into our model a control for size of the male population known to be most involved in criminal activity—those 16 to 29 years of age.[4]

[2] Given the different foci of the Cantor and Land (1985) and Cohen and Land (1987) studies, their results might well be expected to differ. The former attends to short-term fluctuations in annual rates of crime, while the latter assesses long-term relationships. Our intent in comparing these and other diverse studies is to explore the critical unemployment-crime literature that establishes the theoretical context undergirding our hypothesis. To the extent that a given analytical approach resembles our own (as does Cantor and Land's [1985] research), we give it primacy in hypothesis construction.

[3] In a series of substantive and statistical checks, we later substitute several indicators of police expenditures in place of the prison-population rate measure. Except for the use of a measure of deflated (1980 dollars) police expenditures, these estimates are plagued by collinearity and autocorrelation. Coefficients indexing the relationship between the three index crimes and deflated police expenditures are negative, but are not statistically significant (data not presented).

[4] It is important to note that this age group is not

The second variable is criminal opportunity. In brief, criminologists hypothesize that the extent to which members of a population exercise guardianship over their property and limit their exposure to violent crime situations influences that population's crime rates (Cohen and Felson 1979b). Recently, Miethe, Stafford, and Long (1987) report that routine activities and lifestyle variables (i.e., exposure-to-crime variables) are relatively strongly related to risk of property crime but not to risk of violent crime (see also Messner and Blau 1987). Given this finding, we later enter a measure of criminal opportunity into our basic model to tap the aggregate degree of exposure to potential crime situations.

DATA AND MEASUREMENT

We use time-series data to test the above model. Available data for the variables of interest focus our analyses on the years 1948 to 1985, a period of considerable fluctuation in crime rates in the United States.[5]

Crime rate refers, for the purposes of the present study, to *Uniform Crime Reports* (UCR) estimates of homicide, robbery, and burglary rates (per 100,000 persons). These crimes generally have been committed by younger males (the group most troubled by unemployment in the post-World War II U.S.) and represent the types of crimes expected to be influenced by negative economic conditions: burglary as a prototypical crime of property acquisition; robbery as an offense in which actual or threatened violence is used to acquire material goods; and homicide as a crime of frustration and anger

hypothetically occasioned by economic hard times and sometimes accompanying robbery and burglary.

Data for the years 1948 to 1959 derive from the Office of Management and Budget's (1974) revised UCR estimates; those for 1960 to 1985 are taken directly from the UCR (Federal Bureau of Investigation [1961 through 1986]). Although problems with these statistics are well documented (Chilton 1980; O'Brien 1985), they represent the only available source of long-term, national-level crime data.

Because our theoretical interest is the influence of changes in macroeconomic and social-control policies on the changing incidence of crime, the three dependent variables are expressed as first-differences (i.e., change scores). Since the reported rates of homicide, robbery, and burglary exhibit strong secular trends during the period under study, this specification has important statistical properties as well. A first-difference detrends the time-series by eliminating the secular influence in the reported levels of crime.[6] Consequently, we are able to examine the effects of the cyclical economic and social-policy factors hypothesized to affect fluctuations in the reported crime rates (see McCleary and Hay 1980, pp. 31–36; Cantor and Land 1985, pp. 323–26).

The strong secular influence in reported crime rates is shown in the top row of graphs in Figure 2. The middle row of graphs in Figure 2, which depicts the first-differenced time-series (i.e., the change in reported crime rates), demonstrates that the transformed measures effectively remove the secular trend in the original data. While this specification appears to meet the zero-mean criterion of stochastic regression, these graphs also suggest that the required assumption of constant variance may be violated (see Ostrom 1978). With respect to the change scores (middle row, Figure 2), the later years of the series exhibit more year-to-year variability than do the earlier years. Consequently, we follow the lead of Cantor and Land (1985, p. 324) and log-transform and first-difference the original (i.e., level) crime rates (also see McCleary and Hay 1980, pp. 48–53). Graphs of the

---

only high in rates of criminal behavior, but in rates of victimization as well. As Cohen and Land (1987, p. 174) point out in their assessment of the link between age structure and crime rates, the effect of offender behavior generally is indistinguishable from the effect of victim behavior. A more restrictive measure of the male population, the proportion 16 to 24 years of age, was tested in preliminary equations. For homicide and robbery, no significant difference was found between the two age measures. For burglary, however, a statistically significant difference between the coefficients favored the 16–29 year old measure.

[5] Time-series data on all of the hypothesized variables extend back to 1947. However, the use of change scores necessitates the loss of the initial observation. Therefore, the analyses begin with the year 1948.

---

[6] We also experimented with second-order differences to discern whether a quadratic trend was operable. We found no evidence of such a trend.
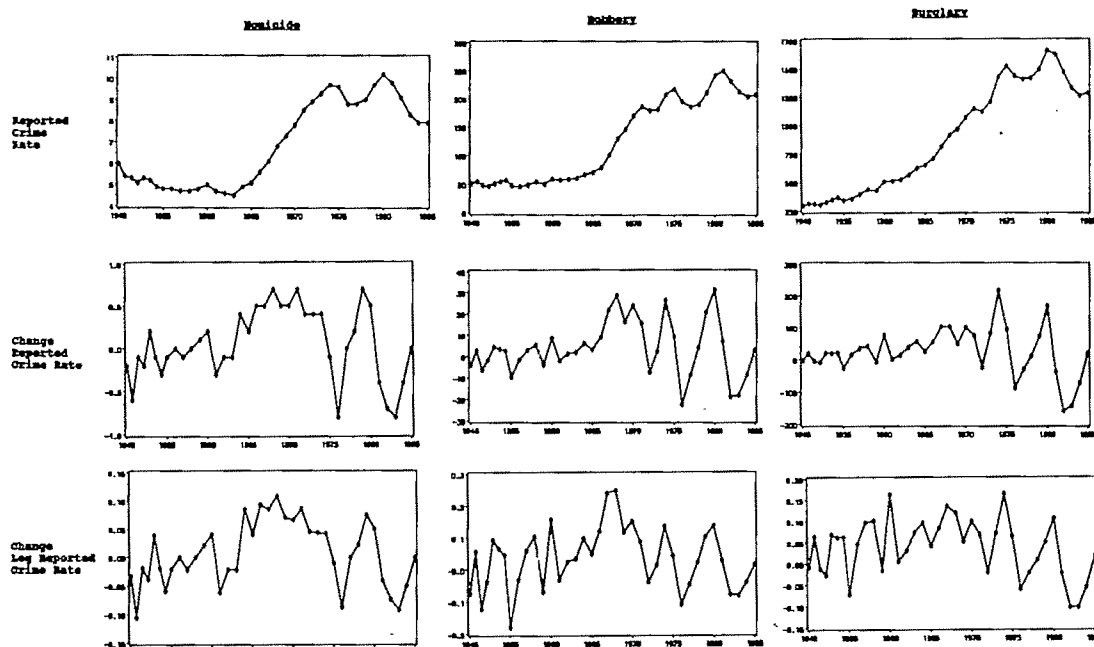
Fig. 2. Level, Change, and Change Log Crime Rates, 1948–85

log-transformed, first-differenced rates of homicide, robbery, and burglary are presented across the bottom row of Figure 2. These graphs suggest that this procedure effectively transforms the data and corrects for the instability, allowing us to meet the variance-stationarity assumption. Given the presence of this instability in the first-difference series, we subsequently model two sets of equations using both the first-difference and log-transformed, first-difference series.

We focus on the *male unemployment rate* because the crimes of interest here characteristically are committed by males (Sheley 1985, pp. 144–49).[7] These data are taken from the *Handbook of Labor Statistics* (Bureau of Labor Statistics 1985) and the *Statistical Abstract of the United States, 1987* (Bureau of the Census 1986). Those regarding *inflation* are found in the *Historical Statistics of the United States* (Bureau of the Census 1975) and the *Statistical Abstract of the United States, 1987* (Bureau of the Census 1986).

Relief statistics derive from data published by the Bureau of Economic Analysis (1976,

---

[7] Age-specific male unemployment measures were tested in place of the aggregate male unemployment variable in preliminary equations. Having found no statistically significant differences across age groups, we use the aggregate indicator.

1977, 1979, 1982, 1983, 1986). This measure represents the sum of public spending on aid to families with dependent children (AFDC), categorical public assistance benefits (i.e., aid to the indigent, blind, and disabled), and other (not elsewhere classified) direct relief.[8]

Prison population information is taken from U.S. Department of Justice (1986a,b) figures. This measure includes both federal-

---

[8] Throughout the analysis, public relief, alternative indicators of public welfare (i.e., including food stamps and other welfare programs), and other spending measures (i.e., police expenditures [see note 3]), are expressed in real (i.e., deflated 1980) dollars. This specification expresses the actual level of resource commitment and minimizes collinearity with the inflation term. In preliminary analyses, the influence of six sequentially more expansive alternative indicators of social welfare spending was explored. These variables ranged from the addition of spending on food stamps to all transfer payments. With greater programmatic aggregation, the postulated relationship to crime proved increasingly precarious: the more aggregate measures exhibited a countertheoretical pattern of nonsignificant positive coefficients. Given that increasing aggregation includes increasingly middle-class benefits (see Devine and Canak 1986), we maintain that the more narrow relief measure is the theoretically and empirically appropriate indicator of placative social control.

and state-held prisoners and is expressed as a rate per 100,000 persons in the general population.[9]

To structure a control variable for age composition of the population, we employ statistics concerning proportion of the *male population 16–29 years old*. These are available in the annual Bureau of the Census's *Current Population Reports,* Series P-25 (1954, 1963, 1974, 1978, 1980, 1984, 1985).

To measure *criminal opportunity* (victim-target availability), Cohen and Felson (1979b) have developed an index of household exposure to risk of residential property crime (for our purposes, burglary). Following their lead, we compute the same measure using data published by the Bureau of the Census (1983, 1986) and the Bureau of Labor Statistics (1985). The index (called a "residential population density ratio" by its authors) divides the sum of the number of female labor-force participants with husband present in the household and the number of nonhusband-wife households by the total number of households. The index taps "household exposure to criminal risk from decreased guardianship either because people live alone or with unrelated persons or because labor force participation removes both husbands and wives from their households during working hours" (Cohen and Land 1987, p. 176).

Because it indicates the number of persons at home to protect against burglaries, we view this same measure as indicative of the extent to which persons are travelling the streets and, thus, exposing themselves to the risk of robbery. We anticipate positive signs regarding the relation of criminal opportunity to both robbery and burglary. The relevance of this measure to the explanation of homicide rates is less straightforward. To the extent a relation between the two exists, it seems more likely that it would be positive, since the majority of homicides are nonfamilial.

Design

In line with the previous discussion, the hypothesized model is estimated with two sets of equations. Within each set we estimate three equations, one for each of the relevant

index crimes. The first set of three equations employs the first-difference series, while the second set of equations uses the log-transformed, first-difference time-series. The general form of the model is:

$$CRC_{it} = a + b_1 CMU_t + b_2 INF_t + b_3 CREL_t + b_4 CPR_t + e_t \quad (1)$$

where, CRC_{it} denotes the first difference (i.e., $RC_{it} - RC_{it-1}$) in crime rate i in year t; CMU_t is the change in the rate of male unemployment in year t; INF_t denotes the rate of inflation (i.e., change in the consumer price index) in year t; $CREL_t$ is the change in public relief in deflated (1980) dollars in year t; and CPR_t represents the change in the rate of the prison population in year t. In addition, a denotes the intercept, $b_{1...n}$ are estimated coefficients, and e_t is the error term.

Our model appears to contain possible simultaneous relationships between the crimes in question and the two measures of social-control policy. On one hand, we presume that the incidence of crime has a causal influence on the rate of incarceration: increasing crime should prompt authorities to expand their efforts to imprison offenders.[10] On the other hand (and, perhaps, in concert with increased incarceration efforts), higher rates of crime might trigger an expansion in the provision of relief as authorities attempt to reduce criminal motivation and discourage further participation in criminal activities by providing alternative means of support (see Piven and Cloward 1971; Hicks and Swank 1983).

In a series of preliminary estimates, we explored these possible simultaneities using two-stage least-squares estimation (2SLS) procedures. With respect to the potential reciprocity between the three index crimes and public relief, we do not find any supporting empirical evidence (data not presented). However, the crime-incarceration (i.e., prison-population rate) reciprocity hypothesis is supported. Consequently, we report 2SLS estimates.[11]

---

[9] A measure of the number of persons imprisoned can be freely substituted in place of the prison-population rate variable without producing any inferential changes.

[10] We thank an anonymous reviewer for noting an important substantive reason why we should find evidence of this simultaneity as well: incarceration statistics are compiled on the basis of a prisoner census taken on December 31 of year t. Hence, the imprisonment rate or its change in year t will be affected (at least in part) by the crime rate or changes in the rate in year t.

[11] Ordinary least squares (OLS) and generalized

Table 1. Effects of Macroeconomic Conditions and Social Control Policies on Crime Rates, 1948–85 (2SLS)

	Equation						
	(1)	(2)	(3)		(4)	(5)	(6)
	First Difference				First Difference Log		
Dependent Variable	Homicide	Robbery	Burglary		Homicide	Robbery	Burglary
Independent Variable							
Chg. Male	.039[a]	4.80	31.90	Log Chg. Male	.061	.295	.227
Unemployment (*t*)	(.873)	(3.58)**	(4.00)**	Unemployment (*t*)	(1.80)*	(4.48)**	(5.23)**
Inflation (*t*)	.070	2.14	10.90	Log Inflation (*t*)	.037	.059	.025
	(4.02)**	(4.04)**	(3.46)**		(4.28)**	(3.50)**	(2.25)*
Chg. Relief-1980	.015–03[b]	−.005–02	−.017	Log Chg.Relief-1980	−.123	−.321	−.358
dollars (*t*)	(.296)	(.321)	(1.88)*	dollars (*t*)	(1.13)	(1.52)†	(2.58)**
Chg. Prison Rate (*t*)	−.069	−1.86	−13.87	Log Chg.	−1.47	−2.62	−1.90
	(5.38)**	(4.82)**	(6.03)**	Prison Rate (*t*)	(5.36)**	(4.95)**	(5.43)**
Adjusted *R* Square	.450	.466	.426	Adjusted *R* Square	.499	.349	.423
Durbin-Watson	1.33	1.33	1.29	Durbin-Watson	1.85	1.83	2.09
Rho	.216	.234	.248	Rho	−.012	.010	−.101

[a] Metric coefficient (*t*-statistic).
[b] Scientific notation.
† $p \leq .10$ (one-tailed).
* $p \leq .05$ (one-tailed).
** $p \leq .01$ (one-tailed).

Finally, we recognize that the "true" effect of the level of the crime rate in the preceding year (i.e., at *t*-1) may not be controlled adequately by first-differencing. Failure to control for the preceding year's level could bias the other coefficients. Such "specification bias" would be indicated by the presence of an otherwise excluded positive lagged endogenous term (see Bohrnstedt 1969, especially pp. 115–16). Therefore, we re-estimate our reported models by including a term that indexes the lagged-level crime rate (i.e., Y_{it-1}). Though this coefficient is positive in 8 of 18 respecifications, it is significant only once (data not presented). Admittedly, the level of significance does not indicate definitively whether specification bias is operable. However, a more critical criterion is available. With only one exception, inclusion of the lagged dependent variable (independent of its sign or significance) yields virtually no meaningful statistical (hence, inferential) difference. This exception pertains to the third equation in Table 2, in which inclusion of a significant negative Y_{t-1} term results in a loss of significance for the coefficient indexing the burglary-relief relationship.[12]

least squares (GLS) estimates yield virtually the same information as the reported 2SLS estimates.

[12] Tabular results of these respecifications are available on request.

ANALYSIS

Estimates of the hypothesized effects of macroeconomic and social-policy influences on changing rates of homicide, robbery, and burglary for the years 1948–1985 are presented in Table 1. The equations in Table 1 suggest that the hypothesized model fits homicide least well, the only consistently significant substantive relationships being the positive influence of inflation and the negative effect of the rate of incarceration. Neither the predicted positive influence of unemployment nor the anticipated inhibitory influence of relief is empirically demonstrated with statistical consistency, though the coefficient indexing the log-homicide-unemployment relationship does achieve statistical significance (column 4). As for the homicide-relief relationship, neither specification yields a significant finding, while the coefficient indexing the first-difference specification is not even in the predicted direction. As far as homicide is the prototypical crime of passion, however, this general lack of explanatory fit is not altogether surprising or unanticipated.

The hypothesized model appears to offer a better explanation of robbery-rate changes (columns 2 and 5) and is particularly well suited to explaining changes in the rate of burglary (columns 3 and 6). With respect to both robbery and burglary (and under either

Table 2. Influence of the Proportion Population Male 16–29 Years of Age on Crime Rates, 1948–85 (2SLS)

	Equation						
	(1)	(2)	(3)		(4)	(5)	(6)
	First Difference				First Difference Log		
Dependent Variable	Homicide	Robbery	Burglary		Homicide	Robbery	Burglary
Independent Variable							
Chg. Male	.082ᵃ	4.77	33.49	Log Chg. Male	.073	.309	.229
Unemployment (*t*)	(1.80)*	(3.29)**	(3.70)**	Unemployment (*t*)	(2.24)*	(4.39)**	(4.99)**
Inflation (*t*)	.015	1.52	5.07	Log Inflation (*t*)	.015	.033	.019
	(.725)	(2.24)*	(1.20)		(1.53)†	(1.56)†	(1.39)†
Chg. Relief-1980	−.041–03ᵇ	−.097–02	−.022	Log Chg. Relief-1980	−.123	−.319	−.356
dollars (*t*)	(.766)	(.572)	(2.12)*	dollars (*t*)	(1.21)	(1.45)†	(2.48)**
Chg. Prison Rate (*t*)	−.107	−2.41	−18.27	Log Chg.	−1.88	−3.09	−1.99
	(5.84)**	(4.13)**	(5.02)**	Prison Rate (*t*)	(5.99)**	(4.59)**	(4.51)**
Prop. Male Pop.	.179	2.52	20.79	Log Proportion	.200	.232	.049
16–29 (*t*)	(3.33)**	(1.47)†	(1.94)*	Male Pop. 16–29 (*t*)	(3.29)**	(1.78)*	(.573)
Adjusted *R* Square	.475	.435	.331	Adjusted *R* Square	.557	.293	.385
Durbin-Watson	1.76	1.46	1.43	Durbin-Watson	2.36	2.00	2.10
Rho	−.004	.192	.197	Rho	−.211	.030	.089

ᵃ Metric coefficient (*t*-statistic).
ᵇ Scientific notation.
† $p \leq .10$ (one-tailed).
* $p \leq .05$ (one-tailed).
** $p \leq .01$ (one-tailed).

specification), the change in the rate of male unemployment exerts a significant expansionary (i.e., positive) effect. Similarly, the rate of inflation is found to have a significant positive influence. While the rate of incarceration appears to inhibit growth in the incidence of both crimes (as well as homicide), the hypothesized negative impact of public relief consistently pertains only to burglary. Though negative regarding robbery, the effect of relief is not even remotely significant in the first-difference equation (column 2) and it only achieves marginal significance in the log equation (column 5).[13]

Rather than addressing the implications of these initial 2SLS findings at this time, we will wait until the report of the statistical analyses is concluded. To summarize briefly,

[13] Heise (1975) argues that a more "generous" significance level is justified when there are relatively few cases because a low number of cases tends to inflate the standard error of the regression coefficient. Having relatively few cases ($N = 38$), we consider coefficients meeting a .10 level of significance as being "marginally" significant, and worthy of careful consideration. Ultimately, however, we draw our inferences based on the conventional .05 criterion of statistical significance. As such, our estimates and subsequent inferences are, perhaps, overly conservative.

however, these initial equations provide two general insights. First, it appears that the log-transformed specifications (columns 4–6) perform slightly better than do the non-logged equations, despite the mixed record of the reported adjusted R^2s. However, these differences have only a minor substantive impact. Second, results suggest that the hypothesized model best fits burglary and least fits homicide. Finally, across all equations, serial correlation is not a problem (as indicated by the Durbin-Watson and rho-statistics), and explained variance is reasonably high.

As noted earlier, young males are responsible for the overwhelming majority of the crimes in question. Thus, it is necessary to control for the influence of this critical demographic group. In Table 2, we incorporate the measure of this crime-prone population—the size of the male population 16–29 years of age expressed as a proportion of the overall population. In all other respects, these models are isomorphic to the specifications presented in Table 1.

The influence of the proportion-of-the-population indicator reveals a positive effect on each of the index crimes. With respect to homicide, an unequivocally significant pattern is observable, while the effect of the demographic variable on the two property

Table 3. Opportunity Effects on Crime Rates, 1948–85 (2SLS)

Dependent Variables	(1)	(2)	(3)		(4)	(5)	(6)
	First Difference			Equation	First Difference Log		
	Homicide	Robbery	Burglary		Homicide	Robbery	Burglary
Independent Variable							
Chg. Male	.094[a]	5.32	34.63	Log Chg. Male	.075	.315	.231
Unemployment (*t*)	(1.99)*	(3.61)**	(3.72)**	Unemployment (*t*)	(2.30)*	(4.53)**	(5.02)**
Inflation (*t*)	.024	1.42	6.21	Log Inflation (*t*)	.021	.037	.021
	(1.18)	(2.28)*	(1.58)†		(2.26)*	(1.87)*	(1.63)†
Chg. Relief-1980	−.055–03[b]	−.155–02	−.024	Log Chg. Relief-1980	−.130	−.330	−.360
dollars (*t*)	(.990)	(.896)	(2.18)*	dollars (*t*)	(1.26)	(1.52)†	(2.50)**
Chg. Prison Rate (*t*)	−.113	−2.75	−18.83	Log Chg.	−1.81	−3.11	−1.99
	(6.06)**	(4.73)**	(5.13)**	Prison Rate (*t*)	(6.12)**	(4.93)**	(4.77)**
Level Opportunity (*t*)	.032–01	.062	.353	Log Level Opportunity (*t*)	.166	.228	.040
	(3.55)**	(2.22)*	(2.01)*		(3.23)**	(2.09)*	(.553)
Adjusted *R* Square	.434	.421	.298	Adjusted *R* Square	.554	.308	.384
Durbin-Watson	1.75	1.57	1.41	Durbin-Watson	2.32	2.07	2.10
Rho	.025	.141	.224	Rho	−.184	−.056	−.091

[a] Metric coefficient (*t*-statistic).
[b] Scientific notation.
† $p \leq .10$ (one-tailed).
* $p \leq .05$ (one-tailed).
** $p \leq .01$ (one-tailed).

crimes is less straightforward. Under the first-difference specification, the robbery-male population relationship is only marginally significant (column 2), while the log-transformed version (column 5) is significant at the .05 criterion. Alternatively, the burglary-male population relationship holds in the first-difference equation (column 3), but not in the log version (column 6).

Across all columns of Table 2, introducing the proportion male population variable undermines the influence of inflation owing to the collinearity between the two measures.[14] Additionally, introducing the demographic term yields a significant homicide-unemployment coefficient in column 1. Though consistent with the posited theoretical relationship, we view the latter term with some caution, as the overall instability of the homicide-unemployment coefficient suggests that this term may be an artifact of particular equation specification.

In Table 3, we introduce the criminal-

opportunity measure. Procedurally, this entails adding the measure of the level of criminal opportunity to our basic 2SLS model. In substantive terms, opportunity is high when fewer persons are at home. Thus, a positive relationship is anticipated as far as the opportunity for criminal victimization is enhanced.

Table 3 indicates that, with respect to both its direction and pattern of significance, the opportunity measure exerts a significant positive effect (the one exception being the log-burglary-opportunity relationship displayed in column 6) and operates in much the same fashion as the male population variable.[15] Moreover, we do not observe any important differences in the other components of these equations when we compare the results of Tables 2 and 3. In short, including the opportunity indicator has no additional material bearing on the behavior of the other coefficients.

---

[14] The zero-order correlation between the proportion male population 16–29 years old and inflation is .679, while the logs correlate at .707. Excluding the inflation term does not alter the exhibited pattern of influence of the population measure.

[15] Since the opportunity and proportion male population measures are highly correlated ($r =$ .609 and .792, for the first-difference and log-transformed, first-difference time-series, respectively), a similar pattern of results is to be expected. At the same time, this collinearity prevents simultaneous inclusion of both constructs in a single equation.

Discussion

We sought to develop and test a model of economic distress, social-control policy, and crime-rate change that is simple, includes both unemployment and inflation as measures of economic distress, and addresses the dual nature of social control. In testing the model, we introduced controls for the age composition of the population and the degree of criminal opportunity.

Our findings provide mixed support for the model. Importantly, for all three crimes, at least one dimension of economic distress and one dimension of social control are influential, buttressing our position that models of crime-rate change must include both macroeconomic and social-control policy elements. Beyond this, changes in burglary rates appear related to changes in both dimensions of economic distress and to both dimensions of social control. With the exception of the influence of relief (the "carrot"), changes in robbery rates are similarly explicable. Changes in homicide rates seem responsive to inflation and changing rates of incarceration, but are unaffected by relief. Unemployment's effect on homicide rates is less clear. While a positive relationship is exhibited across all specifications, there is no statistically significant relation in the original first-difference equation (Table 1, column 1). With the exception of their dampening effect on inflation (due to collinearity), including the age-structure and criminal-opportunity variables does not alter these basic findings. Moreover, as the literature suggests, age structure and criminal opportunity are positively and significantly related to both homicide and robbery. For burglary, however, the results are ambiguous and surprising. We fully expected the influence of both age and opportunity to be most pronounced for this property crime, yet nonsignificant effects are indicated in the log-burglary equations.

Overall, the findings concerning unemployment and crime rates correspond to those found in most of the literature. The positive signs for unemployment and crime-rate changes and the fact that criminal opportunity has been held constant suggest that unemployment motivates crime. The aggregate nature of our data, however, does not permit us to speak to the issue of who commits crimes during periods of economic distress. We note only that upward shifts in the unemployment rate apparently produce shifts in the density distribution of the population toward the higher end of the criminal-motivation continuum.

Our findings regarding prison-population rate changes also are consistent with recent research and supportive of the deterrence/incapacitation thesis. Given their rare appearance in the literature, little can be noted regarding the status of the inflation and relief findings. Inflation's rather strong showing in the present study suggests that future research must address this particular macroeconomic influence on crime. Similarly, relief's demonstrated relation to burglary rates and its marginal significance in the log robbery equations indicate the need for closer scrutiny of this often-neglected aspect of social control.

As we move from attempts to explain more violent offenses to attempts to explain less violent ones (i.e., from homicide to robbery to burglary), our model clearly gains potency. Arguably, our limited ability to model the dynamics of violent crime reflects the inherent limits of a "social-indicators" approach. Our model (or any other exclusively macro-level model) does not begin to tap the critical intervening psychosocial elements implicated in individual decisions to engage in violent personal crime. Moreover, to the extent that rational action is implicated in criminal behavior, we surmise that a hierarchy of motivational complexity is operable. While this hierarchy is collectively mirrored in the seriousness attributed to various crimes by the public (Rossi, Waite, Bose, and Berk 1974) and, by extension, criminal offenders themselves, it is only fully understood through the integration of individual and interpersonal levels of analysis. In short, burglary represents a minimally nonnormative response to real and/or perceived material deprivation. Robbery, by introducing actual or threatened violence into the resolution of the property problem, involves a substantially greater normative violation. As such, it entails a higher social (and physical) risk. Increased risk of social sanctions requires a heightened and possibly more complex motivational commitment and produces a less predictable response. Finally, relative to its role in dispute settlements, homicide's role in property-problem resolution is fairly rare. To the extent that it is involved in a predatory property offense, however, its social prohibi-

tion escalates radically relative to that of robbery.[16] As a wholly violent response to a property problem, premeditated homicide necessitates a far greater and presumably more complex motivational commitment, one not necessarily responsive to macro-level explanation.

Though speculative, this line of reasoning explains the observed differential impact of public relief on the three crimes we have examined. Echoing Cantor and Land (1985), we have argued that, *ceteris paribus,* increasing economic distress prompts an upward shift in the density distribution of the population along the criminal-motivation continuum. As such, the probability of crossing the threshold of criminality would be higher for those persons previously located at or near the motivational margin of legality than it would for those located lower on the continuum. Similarly, the probability of engaging in more serious (i.e., violent) crimes would be greater for those individuals who were previously sufficiently motivated to pursue lesser criminal activities (i.e., theft offenses) than those who had no previous criminal experience. In any case, for those individuals at or near the margin, the provision of public relief might well constitute the (marginal) difference to the extent it offers at least partial compensation for the financial ground lost through economic hardship. However, for those individuals already so far along the motivational continuum that they are involved in robbery and/or homicide, relief would seem so financially inconsequential that it could not mitigate the effects of economic hardship.

The implications of these findings for theories of criminal behavior are important. Liberal and neo-Marxian theorists long have argued that the crime rates of Western industrial nations reflect the inherent instability of the market economy. With respect to property offenses, our findings clearly do not contradict this claim. Nor do our findings contradict the position of these same scholars that relief may be used to placate those classes thought to possess a high criminal propensity. Further, we find no evidence to support the conservative assertion that social welfare has no influence on crime rates, a

result consistent with the claims of more liberal observers of the system. Finally, and perhaps most importantly, our findings indicate that theories of the macroeconomic determinants of crime must also address the social-control policy issue; research must look beyond the narrow context of the criminal-justice system. Social control is broader in scope, as it is in more micro-level relations. Interests are protected through placative as well as deterrent policies.

REFERENCES

Bohrnstedt, George. 1969. "Observations On the Measurement of Change." Pp. 113–33 in *Sociological Methodology, 1969,* edited by Edgar F. Borgatta. San Francisco: Jossey-Bass.

Brenner, M. Harvey. 1976. "Estimating the Social Costs of National Economic Policy: Implications for Mental and Physical Health and Criminal Aggression." Paper No. 5, Joint Economic Committee, Congress of the United States. Washington, DC: U.S. Government Printing Office.

Cantor, David and Kenneth C. Land. 1985. "Unemployment and Crime Rates in the Post-World War II United States: A Theoretical and Empirical Analysis." *American Sociological Review* 50:317–32.

Chilton, Roland. 1980. "Criminal Statistics in the United States." *Journal of Criminal Law and Criminology* 71:56–67.

Chiricos, Theodore. 1987. "Rates of Crime and Unemployment: An Analysis of Aggregate Research Evidence." *Social Problems* 34:187–212.

Cohen, Jacqueline. 1983. "Incapacitation as a Strategy for Crime Control: Possibilities and Pitfalls." Pp. 1–84 in *Crime and Justice: An Annual Review of Research,* vol. 5, edited by M. Tonry and N. Morris. Chicago: University of Chicago Press.

Cohen, Lawrence E. and Marcus Felson. 1979a. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44:588–607.

———. 1979b. "On Estimating the Social Costs of National Economic Policy: A Critical Examination of the Brenner Study." *Social Indicators Research* 6:251–59.

Cohen, Lawrence E. and Kenneth C. Land. 1987. "Age Structure and Crime: Symmetry Versus Asymmetry and the Projection of Crime Rates Through the 1990s." *American Sociological Review* 52:170–83.

Colvin, Mark and John Pauly. 1983. "A Critique of Criminology: Toward an Integrated Structural-Marxist Theory of Delinquency Production." *American Journal of Sociology* 89:513–51.

---

[16] Indicative of this point, all 37 states with capital punishment statutes include a provision for murder in the commission of a property crime.

Currie, Elliott. 1985. *Confronting Crime: An American Challenge*. New York: Pantheon.

Currie, Elliot and Jerome H. Skolnick. 1984. *America's Problems: Social Issues and Public Policy*. Boston: Little, Brown.

DeFronzo, James. 1983. "Economic Assistance to Impoverished Americans: Relationship to Incidence of Crime." *Criminology* 21:119–36.

Devine, Joel A. and William Canak. 1986. "Redistribution in a Bifurcated Welfare State: Quintile Shares and the U.S. Case." *Social Problems* 33:391–406.

Federal Bureau of Investigation. 1961–1986. *Crime in the United States: Uniform Crime Reports*. Washington, DC: U.S. Government Printing Office.

Foster, Edward. 1981. "Who Loses From Inflation?" *Annals of the American Academy of Political and Social Science* 456:32–45.

Heise, David. 1975. *Causal Analysis*. New York: Wiley.

Hicks, Alexander and Duane Swank. 1983. "Civil Disorder, Relief Mobilization, and AFDC Caseloads: A Reexamination of the Piven and Cloward Thesis." *American Journal of Political Science* 27:695–716.

Hirsch, Fred and John H. Goldthorpe, eds. 1978. *The Political Economy of Inflation*. Cambridge, MA.: Harvard.

Hollister, Robinson and John Palmer. 1972. "The Impact of Inflation on the Poor." Pp. 240–70 in *Redistribution to the Rich and the Poor*, edited by Kenneth Boulding and M. Pfaff. Belmont, CA: Wadsworth.

Land, Kenneth C. and Marcus Felson. 1976. "A General Framework for Building Dynamic Macro Social Indicator Models: Including an Analysis of Changes in Crime Rates." *American Journal of Sociology* 82:565–604.

Lindberg, Leon and Charles S. Maier, eds. 1985. *The Politics of Inflation and Economic Stagnation*. Washington, DC: Brookings.

Long, Sharon K. and Ann D. Witte. 1981. "Current Economic Trends: Implications For Crime and Criminal Justice." Pp. 69–143 in *Crime and Criminal Justice in a Declining Economy*, edited by Kevin N. Wright. Cambridge, MA: Oelgeschlager, Gunn, and Hain.

McCleary, Richard and Richard A. Hay. 1980. *Time Series Analysis For the Social Sciences*. Beverly Hills, CA: Sage.

Messner, Steven F. and Judith R. Blau. 1987. "Routine Leisure Activities and Rates of Crime: A Macro-Level Analysis." *Social Forces* 65:1035–52.

Miethe, Terance D., Mark C. Stafford, and J. Scott Long. 1987. "Social Differentiation in Criminal Victimization: A Test of Routine Activities/Lifestyle Theories." *American Sociological Review* 52:184–94.

Minarik, Joseph. 1979. "Who Wins, Who Loses From Inflation?" *Challenge* (Jan.–Feb.):26–31.

Mirer, Thad W. 1973. "The Effects of Macroeconomic Fluctuations on the Distribution of Income." *Review of Income and Wealth* 19: 385–405.

O'Brien, Robert. 1985. *Crime and Victimization Data*. Beverly Hills, CA.: Sage.

O'Connor, James. 1973. *The Fiscal Crisis of the State*. New York: St. Martin's.

Office of Management and the Budget. 1974. *Social Indicators, 1973*. Washington, DC: U.S. Government Printing Office.

Ostrom, Charles W. 1978. *Time Series Analysis: Regression Techniques*. Beverly Hills, CA: Sage.

Parker, Robert N. and Allan Horwitz. 1986. "Unemployment, Crime, and Imprisonment: A Panel Approach." *Criminology* 24:751–73.

Peretz, Paul. 1985. *The Political Economy of Inflation in the United States*. Chicago: University of Chicago Press.

Piven, Frances Fox and Richard A. Cloward. 1971. *Regulating the Poor: The Functions of Public Welfare*. New York: Vintage.

Rossi, Peter H., Emily Waite, Christine E. Bose, and Richard E. Berk. 1974. "The Seriousness of Crimes: Normative Structure and Individual Differences." *American Sociological Review* 39:224–37.

Sheley, Joseph F. 1985. *America's "Crime Problem."* Belmont, CA: Wadsworth.

Steffensmeier, Darrell J. and Miles D. Harer. 1987. "Is the Crime Rate Really Falling? An 'Aging' U.S. Population and Its Impact on the Nation's Crime, 1980–1984." *Journal of Research in Crime and Delinquency* 24:23–48.

Thornberry, Terence P. and R.L. Christenson. 1984. "Unemployment and Criminal Involvement: An Investigation of Reciprocal Causal Structures." *American Sociological Review* 49:398–411.

U.S. Bureau of the Census, Department of Commerce. 1954. "Estimates of the Population of the United States and of the Components of Change, By Age, Color, and Sex: 1940 to 1950." *Current Population Reports*, Series P-25, vol. 98.

———. 1963. "Estimates of the Population of the United States By Age, Color, and Sex: July 1, 1950 to 1962." *Current Population Reports*, Series P-25, vol. 265.

———. 1974. "Estimates of the Population of the United States By Age, Sex, and Race: April 1, 1960 to July 1, 1973." *Current Population Reports*, Series P-25, vol. 519.

———. 1975. *Historical Statistics of the United States: Colonial Times to 1970*. Washington, DC: U.S. Government Printing Office.

———. 1978. "Estimates of the Population of the United States By Age, Sex, and Race: 1970 to

1977." *Current Population Reports,* Series P-25, vol. 721.

———. 1980. "Estimates of the Population of the United States By Age, Sex, and Race: 1976 to 1979." *Current Population Reports,* Series P-25, vol. 870.

———. 1983. *Statistical Abstract of the United States, 1984.* Washington, DC: U.S. Government Printing Office.

———. 1984. "Projections of the Population of the United States By Age, Sex, and Race: 1983 to 2080." *Current Population Reports,* Series P-25, vol. 952.

———. 1985. "Estimates of the Population of the United States By Age, Sex, and Race: 1980 to 1984." *Current Population Reports,* Series P-25, vol. 965.

———. 1986. *Statistical Abstract of the United States, 1987.* Washington, DC: U.S. Government Printing Office.

U.S. Bureau of Economic Analysis, Department of Commerce. 1976. "Survey of Current Business." (July)

———. 1977. *The National Income and Product Accounts of the United States, 1929–1974,*

Statistical Tables. Washington, DC: U.S. Government Printing Office.

———. 1979. "Survey of Current Business." (July)

———. 1982. "Survey of Current Business." (July)

———. 1983. "Survey of Current Business." (July)

———. 1986. "Survey of Current Business." (July)

U.S. Bureau of Labor Statistics, Department of Labor. 1985. *Handbook of Labor Statistics* (Bulletin 2217). Washington, DC: U.S. Government Printing Office.

U.S. Department of Justice, Bureau of Justice Statistics, 1986a. *Prisoners in State and Federal Institutions on December 31, 1983.* Washington, DC: U.S. Government Printing Office.

———. 1986b. *Prisoners in 1985.* Bulletin NJC-101384. Washington, DC: U.S. Printing Office.

Wilson, James Q. 1985. *Thinking About Crime,* rev. ed. New York: Vintage.

Zedlewski, Edwin. 1983. "Deterrence Findings and Data Sources: A Comparison of the Uniform Crime Reports and the National Crime Surveys." *Journal of Research in Crime and Delinquency* 20:262–76.

THE SOCIAL PRODUCTION OF CRIMINAL HOMICIDE: A COMPARATIVE STUDY OF DISAGGREGATED RATES IN AMERICAN CITIES*

KIRK R. WILLIAMS   ROBERT L. FLEWELLING
University of New Hampshire

The growing research on comparative studies of homicide in the United States reveals significant methodological advances but inconsistent findings. A major goal is to identify sources of inconsistency and accumulate more valid and reliable results. This analysis empirically examines a major problem with most previous comparative studies—the failure to disaggregate the overall homicide rate into more refined and conceptually meaningful categories of homicide. A theoretically integrated model is presented that guides the calculation of disaggregated rates and the derivation of hypothesized relationships. Using data from the Comparative Homicide File (CHF), the analysis shows that indicators of resource deprivation and social disintegration tend to have significant effects across subtypes of homicide, although the magnitude of the effects varies, while indicators of violent cultural orientation are confined to homicides resulting from interpersonal conflicts. The implications of the results for comparative research on homicide are discussed.

Comparative research on homicide has become an established tradition in the sociology of crime. Some researchers have accounted for variation in the homicide rate among nations in the contemporary world (e.g., Krahn, Hartnagel, and Gartrell 1986; Avison and Loring 1986). Others have done so across cities, metropolitan areas, or states in the U.S. (e.g., Loftin and Parker 1985; Messner 1982; Loftin and Hill 1974). However, explaining variation statistically has not been

the primary purpose of comparative studies. To the contrary, most researchers have specified theoretical models, guided by numerous perspectives on the social production of lethal violence. Nonetheless, this line of work has been heavily influenced by the subculture-of-violence thesis.

Drawing from the classic work of Wolfgang and Ferracuti (1967), theorists have argued that due to the distinctive historical circumstances of the South, southerners have a cultural orientation that encourages the use of violence to resolve interpersonal conflicts (e.g., Hackney 1969; Gastil 1971). Accordingly, investigators have used southern location or origin as proxies, and they have used the "regional culture of violence" to explain the persistence of a high homicide rate in the South, when other determinants are statistically controlled (e.g., Messner 1983a, 1983b; Doerner 1983; Huff-Corzine, Corzine, and Moore 1986). Similar arguments have been developed to explain the persistence of a high homicide rate among blacks (e.g., Messner 1983a; Sampson 1985; but see contending arguments in Hawkins 1986).

Other investigators have questioned the regional and racial pattern attributed to violent cultures and have doubted whether variation in the homicide rate is primarily a consequence of differential exposure to such cultures. Instead, they see violent behavior as

an adaptive response to the structural conditions under which people live (e.g., Humphries and Wallace 1980; Blau and Blau 1982). Accordingly, investigators have argued that variation in the homicide rate is a function of other social and economic factors that vary by race or region in the United States. The empirical work of these investigators has emphasized measures of economic deprivation, such as inequality and poverty (e.g., Blau and Blau 1982; Williams 1984), or various indicators of social disintegration, such as divorce or mobility (e.g., Blau and Blau 1982; Crutchfield, Geerken, and Gove 1982).

A detailed review of the findings is beyond the scope of this paper. Suffice it to say that they have been inconsistent, with virtually no variable, either proxies for violent cultural orientation or indicators of economic deprivation, demonstrating a robust effect across studies.

Accounting for these inconsistencies should clearly be a major objective for further comparative research. A comprehensive analysis that systematically addresses all possible sources of the problem would be ideal. Realistically, the objective will be accomplished incrementally, with independent investigators focusing on specific issues. We suggest that the process begin with the "dependent variable" (i.e., the homicide rate) and move on to other issues.

The present analysis focuses on a major problem with previous comparative research—the tendency to analyze the homicide rate as if it were a homogeneous category of lethal violence. Theoretical perspectives in this research tradition (i.e., subculture of violence, economic deprivation, and social disintegration) are used to guide the disaggregation of the total rate into conceptually meaningful subtypes of homicide. We then report an analysis that shows that determinants of homicide are differentially associated

with these rates in ways anticipated by the perspectives.

THEORETICAL FRAMEWORK

Previous comparative studies, with some exceptions (e.g., Parker and Smith 1979; Smith and Parker 1980; Sampson 1985), have analyzed variation in the total homicide rate. Use of the total rate, however, can mask or imprecisely reveal empirical relationships indicative of a differential causal process operating in the social production of criminal homicide. Differential effects that are relevant to particular theoretical perspectives can be properly estimated only if disaggregated homicide rates are used. Disaggregation, therefore, should be guided by the theoretical focus of the research problem. The objective of the present study is to disaggregate the total rate into conceptually meaningful subtypes of homicide that fit the theoretical framework of previous comparative studies.

The perspectives guiding these studies, however, do not apply to all forms of homicide. For example, a violent cultural orientation is more likely to produce deaths that spring from heated arguments among intimates than to produce serial killings, contract killings, or random sniper attacks. Figure 1 helps explain the implications of this point. The model includes three theoretical constructs, shown as exogenous variables, that have been viewed as dominant sources of homicide in previous comparative studies. Two of these constructs pertain to objective conditions of living: social disintegration and resource deprivation. Social disintegration is defined as the disruption of social relationships. Resource deprivation refers to the lack of sufficient means to sustain a healthy life. The third exogenous variable, violent cultural orientation, pertains to a more subjective condition of living and refers to shared beliefs that encourage the threat or use of force to
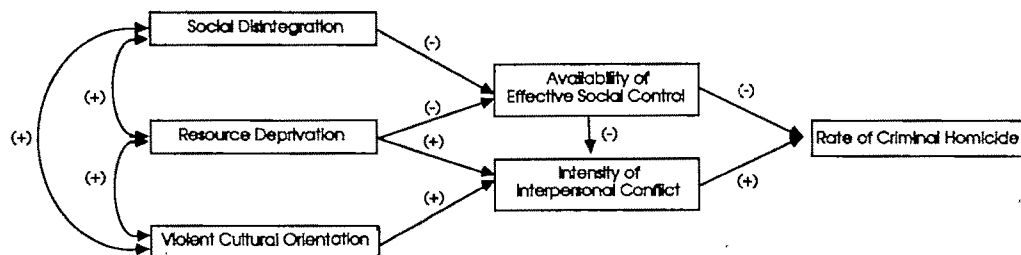
Fig. 1. A Generic Model of the Social Production of Criminal Homicide.

achieve goals or resolve interpersonal conflicts.

THEORETICAL RELATIONSHIPS

The theoretical relationships diagrammed in Figure 1 represent a parsimonious recursive system that simplifies the task of describing the framework, even though a more complicated set of relationships may be involved in this set of constructs. We do not address this issue; rather, a simplified discussion is presented to establish the theoretical context for analyzing disaggregated rates.

Violent cultural orientation and conflict. The rationale for the relationship between a violent cultural orientation and the rate of criminal homicide can be summarized briefly. Antagonistic interactions (i.e., intense interpersonal conflicts) are more accepted among some groups or within some regions of society. People in such groups or regions are more likely to endorse the use of force in settling quarrels or in simply "getting one's way." Consequently, the more a violent cultural orientation penetrates social relationships, the greater the intensity of interpersonal conflicts and the greater the likelihood that some of those conflicts will result in criminal homicide.

Resource deprivation and conflict. The rationale for the relationship between resource deprivation and criminal homicide is not as obvious. Investigators have rarely explained why such deprivation should be positively associated with the homicide rate. The studies by the Blau's (1982) and Messner (1982, 1983a, 1983b) are exceptions, but their arguments often entail obscure implicit assumptions and thus remain debatable. Nonetheless, it is reasonable to assume that when people live under conditions of extreme scarcity, the struggle for survival is intensified. Such conditions are often accompanied by a host of agitating psychological manifestations, ranging from a deep sense of powerlessness and brutalization to anger, anxiety, and alienation. Such manifestations can provoke physical aggression in conflict situations.

We recognize that not all people living in deprivation are violent; people deal with deprivation differently. Since the objective is to account for city-to-city variation in disaggregated homicide rates, it is sufficient to assume that in the aggregate, resource deprivation, like a violent culture orientation, increases the likelihood that lethal means will be used in intense interpersonal conflicts.

Resource deprivation and social control. Notice that another causal connection between resource deprivation and criminal homicide is delineated in Figure 1. Although physical aggression is a method of dealing with conflict situations that is available to virtually all people, it is used as a last resort, that is, when other nonviolent methods of conflict resolution are unavailable or ineffective.

Thus, criminal homicide is more likely when effective social control is less available. Social control is defined as attempts by one person to change the behavior of another person through some form (real or perceived) of third-party intervention (Gibbs 1981). The third party can range from a personal friend to the police, or, put more generally, from informal to formal agents of social control. With this distinction in mind, the argument can be restated: The lower the availability of effective methods of social control, the greater the fatality rate of interpersonal conflicts. In short, homicides that spring from interpersonal conflicts are often the result of failures in social control.

Failures can be due to the ineffectiveness or the unavailability of social control. According to Figure 1, resource deprivation tends to diminish both aspects of social control: if one cannot buy food, clothing, or shelter, one cannot buy "help," be it counseling, legal services, or "a way out." Although public institutions of social control, such as the criminal-justice system, are not necessarily mobilized by one's ability to "pay," the "benign-neglect" thesis suggests those institutions are not responsive to the troubles of the poor and minorities (e.g., Liska, Chamblin, and Reed 1985; Hawkins 1986).

Social disintegration and social control. The rationale for the relationship between social disintegration and criminal homicide can be traced at least to Durkheim, who argued that integrated communities have lower rates of crime. Early work in the Chicago School is also relevant here. As Shaw and McKay (1972) argued, a breakdown of community integration can weaken informal and institutional forms of social control, thus increasing the likelihood of criminal activity (see Bursik 1984 for a recent analysis).

Nonetheless, investigators have applied

these theories to comparative research on homicide. For example, Crutchfield et al. (1982, p. 468) contend that "an integrated social system provides 1) a high degree of consensus in norms, values, and goals; 2) cohesiveness, or social solidarity; and 3) a sense of belonging or 'we feeling' among persons living in the community in question." They argue that factors which impede the maintenance of an integrated social system create a social context conducive to homicide. We would add that a diminution of consensus, cohesiveness, and belongingness is accompanied by a withering of social controls that can create a social environment in which conflicts of values and interests are more likely, thus increasing the probability of violence.

Also, the availability of effective social control may reduce the fatality rate of interpersonal conflict in the case of third-party interventions that allow people, for example, to escape violent situations before they become fatal, such as shelters for battered women (e.g., Browne and Williams 1987). To the extent such interventions are less available or less effective as a result of social disintegration, its impact on the rate of criminal homicide operates apart from the intensity of interpersonal conflict.

The Basis of Disaggregation

The generic model provides an integrated set of theoretical relationships from which expectations about effects can be derived. However, the discussion of the model suggests that expected effects will vary depending on the form of homicide analyzed. The general category of homicide should be disaggregated on the basis of those characteristics that best differentiate the interpersonal and situational contexts in which lethal incidents take place. After all, the exogenous theoretical constructs of the model are asserted to exert a strong impact on the nature of person-to-person interactions in those contexts. Specifically, conflict is more intense, and effective nonviolent means of resolution are less available.

It is possible to disaggregate the general category of homicide in a way that identifies whether killings resulted from conflict situations (e.g., arguments, fights, etc.) and the degree of intimacy between the persons involved. Stated more generally, two criteria guide the calculation of disaggregated rates:

(1) the nature of the precipitating incidents that result in criminal homicide; and (2) the nature of the relationship between victims and offenders.

For our purposes, incidents of interpersonal conflict are most relevant, compared to other events such as felony homicides or contract killings. Included in conflict homicide are those incidents that clearly resulted from a person-to-person argument, quarrel, or fight. Undoubtedly, some forms of homicide in the "other" category may stem from similar circumstances, such as a gambling homicide or a rape homicide if the perpetrator and victim are intimates. Such refinements should be examined in future research.

The victim/offender relationship is an important basis of disaggregation because most homicides occur in the context of the normal, routine patterns of interaction in everyday life. Consistent with previous studies, the degree of intimacy in the relationship between victims and offenders is used as an indicator of the regularity of interaction (e.g., Riedel and Zahn 1985; Loftin and Parker 1985).

On the basis of these two defining characteristics of homicide, rates of homicide based on the cross-classification of conflict and other homicides by family members, acquaintances, and strangers are calculated for a sample of American cities of 100,000 population or more during 1980–84.

Derived Expectations

Expectations can be derived from the theoretical model about the effects of these indicators on the disaggregated rates. Recall that a violent cultural orientation is asserted to escalate the intensity of interpersonal conflict. Hence, indicators of this construct should have greater effects on conflict compared to other homicide rates. Additionally, the effects of such indicators should be greater for family and acquaintance homicides than stranger homicides. This is so because interpersonal conflicts are likely to be more common and more intense among those who regularly interact with one another. In other words, the greater the intimacy in social relationships, the greater the opportunity for antagonistic (often expressive) encounters, and the greater the likelihood of criminal homicide.

The same argument also applies to resource deprivation, but derivations are more compli-

cated because it is not related to criminal homicide solely through intensified interpersonal conflicts. It is also asserted to increase the rate of homicide by reducing the availability of effective means of nonviolent conflict resolution (i.e., social control). Therefore, the effects of indicators of resource deprivation should be more pervasive (i.e., hold across a greater variety of types of homicide) than those for a violent cultural orientation.

Indicators of social disintegration should also have more pervasive effects on the disaggregated homicide rates because they are asserted to decrease the availability of effective social control, which, in turn, increases the rate of criminal homicide either directly or indirectly through the intensification of interpersonal conflict. However, if the primary effect of social disintegration is to reduce the capability of a community to regulate or protect *any* of its members, then estimated effects for rates of other homicide involving less intimate victims and offenders could be as great or greater than those for rates of conflict homicides involving intimates. This expectation, moreover, is consistent with the theoretical imagery found in the history of the social disintegration perspective from Durkheim to the present (i.e., higher rates of deviance in more anonymous and unstable communities).

DATA, RATES, AND INDICATORS

The homicide data analyzed were obtained from the Supplementary Homicide Report (SHR), collected by the Federal Bureau of Investigation (FBI) as a part of its Uniform Crime Reporting (UCR) program. Detailed information from the SHR is not published in the FBI's annual *Crime in the United States*. The complete SHR files can be acquired directly from the FBI, although our project will produce public use computer files of homicide rates for states, metropolitan areas, and cities to be distributed by the National Institute of Justice. Those files will also include numerous variables pertaining to the social, economic, and demographic characteristics of these units. The entire data set is referred to as the *Comparative Homicide File* (CHF).

Homicide is subdivided into three categories: murder and nonnegligent manslaughter, negligent manslaughter, and justifiable homicide. This research focuses exclusively on

murder and nonnegligent manslaughter because previous comparative studies have confined their analyses to this type. Furthermore, it represents that most common form, constituting 94.7 percent of all known incidents in the United States during 1980–84. The term *homicide* will be used hereafter to denote incidents of murder and nonnegligent manslaughter.

Sample of Incidents

Among these incidents of homicide, the sample is restricted to one-on-one cases. This is consistent with our theoretical emphasis on interpersonal conflict. In addition, previous descriptive analyses have shown that the age, sex, and race-ethnic composition, as well as the situational circumstances of incidents involving multiple offenders and/or victims are different from one-on-one incidents (e.g., Block 1985). The implication is that the structural and cultural dynamics producing such incidents differ from one-on-one homicides. These two kinds of incidents should be analyzed separately until the crucial points of similarity and difference are clearly identified. Besides, incidents involving multiple offenders and/or victims are relatively infrequent, representing about 11 percent of all known homicides in the United States during 1980–84.

Furthermore, technical aspects of the data preclude including incidents involving multiple offenders and/or victims; specifically, the relationship-specific rates are based on family, acquaintance, and stranger relations. The SHR data provide information only on the relationship between the first victim and each offender in multiple-victim incidents. Consequently, relationships based on combinations of second and subsequent victims cannot be determined.

Rate Calculation

The incidents of homicide used in rate calculations cover the entire 1980–84 period, not individual years. This procedure was used to reduce the influence of random abberations in year-to-year estimates, in addition to the possible unreliability of rates based on low frequencies. The focus of the theoretical model guiding this analysis, moreover, is on underlying empirical patterns, not short-term fluctuations. The calculation of rates based on

victim/offender relationships and precipitating incidents is complicated by two problems with the SHR.

First, reporting agencies occasionally fail to submit monthly SHR forms to the UCR office. The result is that rates calculated from SHR data will be underestimated in areas with nonreporting. Nonetheless, the UCR office provides "estimated totals" of homicide victims in *Crime in the United States.* So, the extent to which the number of victims in the SHR underestimates the total number of victims in the UCR can be estimated for any given year. A weighting procedure for rate calculations that compensates for nonreporting in the SHR was devised (for a full discussion, see Williams and Flewelling 1987).

Second, among the homicide incidents reported, information on the victim/offender relationship is often missing. For example, data are missing for approximately 25 percent of the incidents, on the average, in the sample of cities analyzed. Such missing data can result in the underestimation of relationship-specific rates and in biased parameter estimates. This problem can be addressed by using an adjustment procedure that incorporates such missing data in the rate calculations. The general strategy of the adjustment procedure is to extrapolate the characteristics of the known cases to those with missing information. Within this general strategy, the adjustments are determined and applied separately on the basis of the nature of precipitating incidents. This refinement of the general strategy takes advantage of what *is* known about the circumstances of incidents with missing information, yielding more accurate estimates of family, acquaintance, and stranger homicide rates (again, see Williams and Flewelling 1987 for a detailed description of the procedure).

Using these weighting and adjustment procedures, relationship-by-event-specific rates are calculated as follows:

Specific
homicide rate $= ((I/P) \times 100,000)/5,$
where $I =$ the total number of weighted and adjusted incidents of murder and nonnegligent manslaughter of a specific type (e.g., incidents involving strangers and conflicts), and
$P =$ the total population of the city.

The division by five indicates that the rates

are calculated over the 1980–84 period, and then expressed on a per-year basis.

Indicators of the Theoretical Model

No effort is made to include all possibly relevant indicators of the theoretical model or to address the many theoretical and methodological issues in comparative research on homicide. The objective is merely to determine whether the theoretical model guiding the disaggregation of the total homicide rate accounts for differential effects. The analysis, therefore, is viewed as a preliminary step in unraveling the social production of specific forms of criminal homicide. A sampling of indicators is used to reflect the basic theoretical constructs employed in previous comparative studies.

Deprivation, disintegration, and control variables. The percent of families below the U.S. Social Security Administration's poverty line (i.e., percent poor) is used as an indicator of resource deprivation (U.S. Bureau of the Census 1980), and the number of divorces per 1,000 population age 15 to 59 (i.e., the divorce rate, U.S. Bureau of the Census 1982) is used as an indicator of social disintegration. The number of divorces is not available in the *State and Metropolitan Area Data Book* at the city level, so rates were calculated for the metropolitan areas in which the cities are located and then assigned to those cities.

Regional location is measured by creating a Confederate south/nonsouth dummy variable. Two additional demographic variables are used: racial composition (i.e., percent black, U.S. Bureau of the Census 1980) and population per square mile (i.e., population density, U.S. Bureau of the Census 1983).

Regional location and racial composition are not used here as proxies for a violent cultural orientation because the significant effects of region or race are "cultural" effects only if one assumes that (1) a violent culture does, in fact, exist in the South or among blacks; (2) it is a primary source of violence among such people; and (3) all other noncultural sources of regional and racial differences in the homicide rate are accurately measured and included in the model estimated. The model estimated does not justify such assumptions, so these variables are not given special theoretical significance.

An alternative indicator of violent orientations. Many have called for more innovative

strategies of measuring violent cultural orientations. Some interesting leads can be found in the research of Archer and Gartner (1984) or Bowers (1984). They offer examples of how the official justification of lethal violence (e.g., war, the imposition of the death penalty) can promote homicide. They also discuss various learning processes that produce this result (e.g., role modeling, devaluation of the victim, loosening of normative constraints). However, all of them involve internalizing the message that problem situations and problem people often justify the intentional destruction of human life.

Drawing from this work, we explore an alternative indicator of violent cultural orientation. The underlying assumption is that an official justification of intentional killing expresses the cultural approval of violence. Such a ruling involves an interpretive process (i.e., it is a matter of definition); and while it undoubtedly is constrained by legal criteria of decision-making, the influence of values is inevitable, as suggested by research on discrimination and the death penalty (e.g., Bowers 1984). Hence, the greater the approval of violence to resolve conflicts of interest, the greater the likelihood that intentional killings will be interpreted by officials as justifiable.

The SHR includes data on justifiable homicides, classified as such by reporting police agencies and subdivided as incidents perpetrated by either police or citizens. The total number of justifiable homicides for the 1980–84 period is used to calculate the indicator of violent cultural orientation.

It is expressed as the number of justifiable homicides per 100 criminal homicides. The reason for the ratio is that justifiable homicides are likely to be more frequent in some cities simply because residents are confronted with more life-threatening situations than elsewhere. Assuming that the incidence of criminal homicide is highly correlated with the frequency of such situations, the calculation procedure "controls" for this possibility. The ratio estimates the flexibility of interpretive standards relative to the volume of criminal homicide.

Transformations. The distributions of the homicide rates are skewed, as are those of some of the independent variables (i.e., divorce rate, justifiable homicide ratio, population density). Furthermore, previous comparative research has shown that the relationship between the total homicide rate and the percent black, as well as the percent poor, is nonlinear (e.g., Messner 1983a; Williams 1984). Therefore, the rates and the independent variables are logarithmically transformed (base 10).

EMPIRICAL FINDINGS

The diverse composition of homicide is illustrated by the distribution of incidents across the victim/offender relationship and precipitating incident categories in Table 1. Observe that an estimated 47,739 incidents occurred in American cities of 100,000 population or more during 1980–84, which is about 54 percent of all such incidents in the United States for this period. Approximately two-thirds of the incidents involved either family members or acquaintances. The other third involved strangers, which is significantly greater than for noncity areas, since stranger homicide accounts for about 25 percent of the total incidents in the entire nation. The higher percentage in Table 1 suggests that lethal violence among strangers occurs more frequently in large urban centers. This compositional difference in homicides between urban and rural areas is an example

Table 1. The Distribtuion of Weighted and Adjusted One-on-One Incidents of Murder and Nonnegligent Manslaughter in American Cities (100,000 Population or More, $N = 168$) by Victim/Offender Relationship and Nature of Precipitating Incident, 1980–1984

Victim/Offender Relationship		Precipitating Incident		
		Conflict[a]	Other[b]	Totals
Family	N	4,900	3,532	8,432
	Row %	58.1	41.9	—
	Col %	23.2	13.3	17.7
Acquaintance	N	13,474	10,126	23,600
	Row %	57.1	42.9	—
	Col %	63.9	38.0	49.4
Stranger	N	2,720	12,987	15,707
	Row %	17.3	82.7	—
	Col %	12.9	48.7	32.9
Totals	N	21,094	26,645	47,739
	Row %	44.2	55.8	—

[a] Lovers' triangle, child killed by baby-sitter, brawls involving alcohol or drugs, arguments over money or property, other arguments.

[b] Rape, burglary, larceny, robbery, motor vehicle theft, arson, other felony types and suspected felony, prostitution and commercial vice, other sex offenses, narcotics drug laws, gambling, gangland killings, juvenile gang killings, institutional killings, sniper attackers, other nonfelony types.

of the variability that may exist across city, county, or state samples, and may be one source of the inconsistent findings reported in previous comparative research.

The cross-classification of incidents suggests that their joint distribution offers important information for defining characteristics of homicide. Stranger homicides occur more frequently in other circumstances, while homicides involving family members or acquaintances are more likely to be the consequence of interpersonal conflicts. Although not uniform, the general pattern in Table 1 can be summarized as follows: The greater the intimacy in the relationship between victims and offenders, the greater the likelihood that homicides result from conflict situations; conversely, the greater the relational distance between victims and offenders, the more homicides tend to result from other circumstances (e.g., robberies or other felonies).

Although Table 1 demonstrates a strong relationship, there is variation in the joint distribution that cannot be adequately expressed by a single variable. The conclusion to be drawn from Table 1 is that homicide in large American cities is a heterogeneous phenomenon that can be categorized by two related characteristics: the victim/offender relationship and the nature of the precipitating incident.

Correlations Between Rates

The use of the total homicide rate implies that it is representative of the perpetration rate in its constituent categories. To test this implication, we computed Pearson's Product Moment Correlations between the relationship-by-incident-specific rates and the total homicide rate (see Table 2).

Observe that the correlations between each of the relationship-by-incident-specific rates and the total rate vary in magnitude (from $r = .650$ to $r = .906$). Such variation is even greater for the intercorrelations among the relationship-by-incident-specific rates (from $r = .363$ to $r = .766$). The variability in the correlations suggests that the total rate is not representative of the perpetration rate for these subtypes of homicide. Stated alternatively, the results suggest that estimates of theoretical models could differ depending on the type of rate analyzed.

Table 2. Correlation Matrix for Victim/Offender Relationship by Precipitating Incident-Specific Rates and the Total Homicide Rate for American Cities (100,000 Population or More, $N = 168$), 1980–1984[a]

	(1)	(2)	(3)	(4)	(5)	(6)
1. Family conflict						
2. Family other	.481					
3. Acquaintance conflict	.766	.526				
4. Acquaintance other	.563	.631	.692			
5. Stranger conflict	.587	.371	.515	.363		
6. Stranger other	.629	.452	.658	.596	.672	
7. Total rate	.801	.673	.906	.831	.650	.829

[a] All rates are logarithmically transformed (base 10) to adjust for skewness.

Estimated Effects

To determine whether such variation exists, and whether it conforms to the expectations of the theoretical model, multivariate equations were estimated using Ordinary Least Squares (OLS).

Unstandardized and standardized OLS estimates for the six relationship-by-incident-specific rate equations and the total-rate equation are presented in Table 3. Notice that the estimated effects of all the variables in the total-homicide-rate equation are statistically significant except for location in the Confederate South. These findings suggest that resource deprivation, violent cultural orientation, and social disintegration play an important role in producing city-to-city variation of criminal homicide. However, the estimates reported for the total-rate equation are not consistently sustained in the other equations in Table 3.

The estimated effects of the percent poor remain significant and positive across all equations, but the magnitude of the estimates varies, particularly in comparison to the total-rate equation. That variation does not appear to be systematically associated with either the victim/offender relationship or the nature of the precipitating incident. Nonetheless, this is consistent with the theoretical expectation that resource deprivation has a more pervasive impact on various forms of criminal homicide than do other structural or cultural determinants.

A similar point can be made about the estimated effects of the divorce rate. Estimates are significant and positive in all equations, with the exception of conflict homicides involving strangers. Moreover, the

Table 3. OLS Estimates of the Effects of Indicators of the Theoretical Model on Victim/Offender Relationship by Precipitating Incident-Specific Rates and the Total-Homicide Rate for American Cities (100,000 Population or More, $N = 168$), 1980–84[a]

Indicators		Family Conflict	Family Other	Acquaintance Conflict	Acquaintance Other	Stranger Conflict	Stranger Other	Total Rate
Percent poor	b	.255***	.355***	.542***	.512***	.277***	.407***	.691***
	s.e.	.087	.095	.101	.126	.088	.139	.105
	Beta	.275	.366	.383	.368	.342	.274	.436
Percent black	b	.140***	.055	.254***	.169***	.057	.215***	.259***
	s.e.	.035	.039	.041	.051	.036	.056	.042
	Beta	.378	.150	.449	.304	.175	.362	.409
Justifiable homicide ratio	b	.054*	.010	.089***	−.011	.060**	.075	.088**
	s.e.	.029	.032	.034	.042	.030	.047	.035
	Beta	.108	.021	.118	−.015	.139	.095	.104
Confederate South	b	.035	.052*	.042	.050	.000	−.048	.045
	s.e.	.025	.027	.029	.036	.026	.040	.030
	Beta	.095	.134	.075	.090	.000	−.081	.072
Divorce rate	b	.296***	.341***	.385***	.620***	.170	.566***	.656***
	s.e.	.102	.113	.119	.149	.104	.164	.124
	Beta	.194	.226	.165	.271	.127	.232	.252
Population density	b	−.015	−.028	.035	.166***	.020	.253***	.118**
	s.e.	.039	.043	.045	.057	.040	.062	.047
	Beta	−.028	−.052	.042	.202	.041	.289	.126
Constant		−.370	−.405	−.821	−1.442	−.441	−1.719	−1.131
R^2		.468	.343	.692	.501	.278	.468	.735

[a] All rates and variables are logarithmically transformed (base 10), except Confederate South.

* $p \leq .005$, one-tailed test of statistical significance.

** $p \leq .025$, one-tailed test of statistical significance.

*** $p \leq .005$, one-tailed test of statistical significance.

estimated effects of the divorce rate tend to be greater for other than conflict homicides, and those effects tend to increase as the intimacy of the victim/offender relationship decreases. This pattern suggests that the causal force of social disintegration directly through the availability of effective social control may be greater in the production of criminal homicide than indirectly through the intensification of interpersonal conflict.

The estimated effects of the justifiable-homicide ratio are significant and positive for all forms of conflict homicide, which is consistent with the expectation derived from the theoretical model. These findings also suggest that the ratio is an improvement over previous attempts to measure violent cultural orientation. It deserves further examination in future research.

Furthermore, given the empirical pattern found with the justifiable-homicide ratio, along with the logic of the culture-of-violence thesis, one would expect the percent black and location in the Confederate South to have similar estimate effects if these indicators reflect a violent cultural orientation. How-

ever, regional location has a significant effect only for other (not conflict) homicides involving family members. While the percent black has significant positive effects on conflict homicide involving family members and acquaintances, it also has such effects on other homicides involving acquaintances and strangers. So, whether racial composition reflects violent cultural orientation remains a debatable issue, and the reason why it is associated with disaggregated rates of criminal homicide is unanswered.

Finally, the estimated effects of population density are significant and positive for other homicides involving acquaintances or strangers. The impersonal social climate often associated with densely populated cities appears to be related to more anonymous forms of homicide, not those that erupt out of intense conflicts between intimates.

SUMMARY AND CONCLUSION

Previous comparative studies of criminal homicide have reported inconsistent results. One possible source of such inconsistencies is

the use of the total-homicide rate. To determine if this is so, we tested an integrated theoretical model using homicide rates disaggregated by the victim/offender relationship and the precipitating incident for a sample of American cities in 1980–84. The findings revealed significant evidence of differential effects. All of the indicators except regional location had significant effects on the total-homicide rate, but estimation using the disaggregated rates showed that some of the indicators have rather extensive effects (e.g., percent poor, divorce rate), while the effects of others (e.g., justifiable-homicide ratio, population density) are more limited. Although the effects of the indicators varied, that variation was not random. Rather, the patterns observed were theoretically expected.

The conclusion from these findings is not that analyses of the total-homicide rate are meaningless and should be discontinued. Studying variation in the total rate is relevant for understanding general patterns of homicide in America, irrespective of variation in its constituent categories, just as studying the Dow Jones Industrial Average is relevant for understanding stock-market performance, regardless of changes in the specific securities that comprise it. Social scientists regularly analyze such global indicators for good reasons. Nonetheless, research involving the total-homicide rate should be interpreted cautiously. The failure to find a significant effect of a particular structural, cultural, or demographic variable does not necessarily mean it is irrelevant. Its effect may be unique to a specific form of homicide.

Finding a significant effect on the total rate, moreover, does not necessarily mean that the variable in question is causally relevant for all forms. Such matters can be determined empirically only if disaggregated rates are analyzed, with disaggregation guided by one's substantive or theoretical focus. A more detailed and complete understanding of the social production of homicide can be achieved by analyzing such rates. Furthermore, affirming or refuting theoretical models on the basis of empirical tests using the total rate is premature. The scope of such models should be delimited empirically by testing their applicability to specific types of homicide.

REFERENCES

Archer, Dane, and Rosemary Gartner. 1984. *Violence and Crime in Cross-National Perspective*. New Haven, CT: Yale University Press.

Avison, William R. and Pamela L. Loring. 1986. "Population Diversity and Cross-National Homicide: The Effects of Inequality and Heterogeneity." *Criminology* 24:733–49.

Blau, J.R. and P.M. Blau. 1982. "Metropolitan Structure and Violent Crime." *American Sociological Review* 47:114–128.

Block, Carolyn Rebecca. 1985. "Lethal Violence in Chicago over Seventeen Years: Homicides Known to the Police, 1965–1981." Chicago: Illinois Criminal Justice Information Authority.

Bowers, William J. 1984. *Legal Homicide*. Boston: Northeastern University Press.

Browne, Angela and Kirk R. Williams. 1987. "Resources Availabililty for Women at Risk: Its Relationship to Rates of Female-Perpetrated Partner Homicide." Paper presented at the 1987 meeting of the American Society of Criminology, Montreal, Canada.

Bursik, Robert J., Jr. 1984. "Urban Dynamics and Ecological Studies of Delinquency." *Social Forces* 63:393–413.

Crutchfield, Robert D., Michael R. Geerken, and Walter R. Gove. 1982. "Crime Rate and Social Integration." *Criminology* 20:467–78.

Doerner, William G. 1983. "Why Does Johnny Reb Die When Shot?" The Impact of Medical Resources upon Whodunnit." *Sociological Inquiry* 53:1–15.

Gastil, Raymond D. 1971. "Homicide and a Regional Culture of Violence." *American Sociological Review* 36:412–27.

Gibbs, Jack P. 1981. *Norms, Deviance, and Social Control*. New York: Elsevier

Hackney, Sheldon. 1969. "Southern Violence." Pp. 505–27 in *Violence in America*, edited by Davis Graham and Ted Robert Gurr. New York: Bantam.

Hawkins, Darnell F., ed. 1986. *Homicide Among Black Americans*. Lanham, MD: University Press of America.

Huff-Corzine, Lin, Jay Corzine, and David C. Moore. 1986. "Southern Exposure: Deciphering the South's Influence on Homicide Rates." *Social Forces*. 64:906–24.

Humphries, Drew and Don Wallace. 1980. "Capitalist Accumulation and Urban Crime, 1950–1971." *Social Problems* 28:179–93.

Krahn, Harvey, Timothy F. Hartnagel, and John W. Gartrell. 1986. "Income Inequality and Homicide Rates: Cross-National Data and Criminological Theories." *Criminology* 24:269–95.

Liska, Allen E., Mitchell B. Chamblin, and Mark D. Reed. 1985. "Testing the Economic Production and Conflict Models of Crime Control." *Social Forces* 64:119–38.

Loftin, Colin and Robert H. Hill. 1974. "Regional

Subculture and Homicide." *American Sociological Review* 39:714–24.

Loftin, Colin and Robert Nash Parker. 1985. "An Errors-in-Variable Model of the Effect of Poverty on Urban Homicide Rates." *Criminology* 23:269–85.

Messner, Steven F. 1982. "Poverty, Inequality, and the Urban Homicide Rate." *Criminology* 20:103–14.

———. 1983a. "Regional and Racial Effects on the Urban Homicide Rate: The Subculture of Violence Revisited." *American Journal of Sociology* 88:997–1007.

———. 1983b. "Regional Differences in the Economic Correlates of Urban Homicide Rates." *Criminology* 21:477–88.

Parker, Robert Nash and M. Dwayne Smith. 1979. "Deterrence, Poverty, and Type of Homicide." *American Journal of Sociology* 85:614–24.

Riedel, Marc and Margaret A. Zahn. 1985. *The Nature and Patterns of American Homicide.* Washington, DC: U.S. Department of Justice, National Institute of Justice.

Sampson, Robert J. 1985. "Race and Criminal Violence: A Demographically Disaggregated Analysis of Urban Homicide." *Crime and Delinquency* 31:47–82.

Shaw, Clifford R. and Henry D. McKay. 1972. *Juvenile Delinquency and Urban Areas.* Chicago: The University of Chicago Press.

Smith, M. Dwayne and Robert Nash Parker. 1980.

"Type of homicide and Variation in Regional Rates." *Social Forces* 59:136–47.

Swigert, Victoria Lynn and Ronald A. Farrell. 1976. *Murder, Inequality, and the Law.* Lexington, MA: Lexington Books.

United States Bureau of the Census. 1980. General Social and Economic Characteristics. Washington, DC: U.S. Government Printing Office.

———. 1982. State and Metropolitan Area Data Book. Washington, DC: U.S. Government Printing Office.

———. 1983. County and City Data Book. Washington, DC: U.S. Government Printing Office.

Williams, Kirk R. 1984. "Economic Sources of Homicide: Reestimating the Effects of Poverty and Inequality." *American Sociological Review* 49:283–89.

Williams, Kirk R. and Robert L. Flewelling. 1987. "Family, Acquaintance, and Stranger Homicide: Alternative Procedures for Rate Calculations." *Criminology.* 25:543–560.

Williams, Kirk R. and Murray A. Straus. 1988. *Comparative Homicide File* [MRDF]. Durham, NH: Family Research Laboratory, University of New Hampshire [producer]. Washington, DC: National Institute of Justice [distributor].

Wolfgang, Marvin E., and Franco Ferracuti. 1967. *The Subculture of Violence.* Beverly Hills: Sage Publications.

CROSS-NATIONAL DETERMINANTS OF CHILD HOMICIDE*

ROBERT FIALA   GARY LAFREE
The University of New Mexico

This paper uses cross-national data on child homicide to develop a societal-level model of child abuse. Analysis of hypotheses emerging from four perspectives on abuse suggests that a focus on economic stress is most useful in explaining cross-national variation in child homicide in developed countries. High levels of female labor-force participation,· coupled with low levels of female status and welfare spending, result in high child-homicide rates. No perspective provides hypotheses able to explain cross-national variation in child homicide in less-developed countries. The limitations of the present study are noted and avenues for future empirical and theoretical work are discussed.

Contemporary attempts to understand child abuse have conceptualized it as more than a consequence of individual pathology. Increasing attention is being directed to various aspects of social structure, including stress generated by economic hardship (Gil 1975; Straus, Gelles, and Steinmetz 1980); the breakdown of normative systems under the impact of social change (Gelles and Pedrick-Cornell 1983; Korbin 1987); violence-legitimating cultures (Gil 1975; Straus et al. 1980); and social isolation (Korbin 1981; Garbarino 1987). Such views should be tested by cross-cultural comparative studies. Although an increasing number of case studies have been conducted in various nations, the lack of comparable data has hindered comparative research. This study uses cross-national data on child homicide to help clarify societal characteristics most clearly linked to child abuse. Our results show that a focus on economic stress offers the best explanation of national variation in child-homicide rates. Yet, contrary to most research applying this perspective, our analysis suggests that it is the

economic stress associated with the societal status of women that most effectively explains child homicide.

MEASURING CHILD ABUSE AND HOMICIDE

Recent research on child abuse has moved beyond a focus mostly on the United States, to become international in character (Korbin 1981, 1987; Gelles and Pedrick-Cornell 1983; Gelles 1987). However, the literature still reflects a proliferation of studies within individual countries and almost no comparative cross-national research. While case studies provide details of historical factors operating in specific societies, cross-national analyses offer the range of variation needed to make generalizations about societal characteristics.

The major impediments to such analyses have been conceptual and methodological. First, child abuse is difficult to define. For example, following the influential work of Kempe, Silverman, Steele, Doegemueller, and Silver (1962) on the "battered child syndrome," most researchers presumed that it referred to children who had been deliberately injured by a parent or caretaker. However, this definition has gradually been replaced by much broader conceptions, including not only diagnosable physical and medical symptoms, but also emotional and mental abuse (cf. Gil 1970; Gelles 1974; Korbin 1981).

Definitions of abuse are even more variable in a cross-cultural context. As Korbin (1981, 1987) points out, the main problem with establishing meaningful cross-national definitions of child abuse is that they must ultimately rest on values: what is considered

acceptable treatment in one culture may be viewed as unacceptable or even detrimental in another. For example, Korbin (1981) notes that the common Euro-American practice of making infants sleep in beds separate from their parents is considered uncaring and even abusive in many of the world's societies. In contrast, a diet that is considered evidence of abuse in developed nations may represent the maximum available in impoverished Third World nations.

Obviously, with little cross-national agreement on the definition of abuse, measuring it becomes quite difficult. In the United States, child abuse records have been available only since the 1960s. According to Gelles (1987, p. 13), the only other country that maintains such records is Canada. But even if more nations collected data on child abuse, their validity would be questionable. Burgdorf (1980) estimates that in the United States, which has mandatory child-abuse reporting laws, only about one-third of all instances of child abuse are officially recorded. Other researchers (e.g., Gelles 1975; Newberger, Reed, Daniel, Hyde, and Kotelchuck 1977) argue that child-abuse data on the poor, black, or ethnic minorities are especially prone to reporting variations.

To avoid some of these problems, we concentrate on cross-national variation in child-homicide rates. Such an approach has at least four advantages. First, compared to other types of data on child abuse, homicide data are the most reliable and valid (Steinmetz and Straus 1974; Gelles and Straus 1979). While undoubtedly some instances of child homicide are never identified, missing children are more difficult to conceal. Second, while only a few countries report child-abuse data, child-homicide data are widely available. Third, while definitions of child abuse vary across cultures, child homicide is almost universally outlawed. Finally, because child homicide is often the result of prior episodes involving less serious forms of abuse (Sims and Cameron 1973; Maney and Kedem 1982), the relationship between homicide and other variables should have more general implications for the etiology of child abuse, especially abuse involving violence towards children.

SOCIAL STRUCTURAL PERSPECTIVES ON CHILD ABUSE AND HOMICIDE

Despite conceptual and methodological advan-

tages, little comparative cross-national research has appeared on child homicide. The main exception is that of Christoffel and her associates (Christoffel, Liu, and Stamler 1981; Christoffel and Liu 1983). Using United Nations data, they showed that child- and adult-homicide rates are not highly correlated, a finding they interpret as indicating different etiological processes. However, although Christoffel et al. suggest some explanations for variations in child-homicide rates, they did not directly test them.

In contrast to the paucity of cross-national research on child-homicide rates, a rapidly expanding literature has appeared on the social-structural causes of child abuse. This literature suggests four social-structural perspectives on abuse: economic stress, social disorganization, culture of violence, and social isolation. The first two perspectives are the most prominent, while the other two provide additional hypotheses important for understanding cross-national variation in child homicide.

Economic Stress

Perhaps the most common social-structural explanation of child abuse is that it results from stress caused by economic hardships associated with inequality, poverty, and unemployment (Gil 1975; Gelles and Straus 1979). In general, researchers have argued that economic problems cause child abuse by increasing family stress and/or threatening the male's ability to fulfill his role as family provider (Goode 1971; Steinmetz 1978; Straus et al. 1980).

Although we know of no cross-national tests of this argument for child homicide, studies of overall homicide rates within countries are generally congruent. Thus, a substantial cross-national literature shows a connection between economic inequality and homicide rates (Hansmann and Quigley 1982; Messner 1982; LaFree and Kick 1986), and intra-societal research reveals that child abuse is more common in lower- than upper-income households (Gil 1970; Gelles 1974; Straus et al. 1980). Relatedly, research has indicated that unemployment is associated with abuse. For example, Gil (1970) found that 48 percent of the fathers in his U.S. sample of child abusers had experienced unemployment during the year preceding the abuse—a finding later confirmed in a reanalysis of a matched sample of nonabusing families (Light 1974).

Following Goode's (1971) influential work, much of the literature on the effect of economic stress on child abuse has focused on its impact on adult males. However, the fact that women typically assume primary responsibility for child rearing suggests equal, if not more, attention be given to the status of women in society. Most literature touching on this issue (e.g., Williams 1980; Breines and Gordon 1983) suggests that societies in which the role structure limits opportunities for women and increases the difficulty of their tasks should experience increased household stress, with a concomitant increase in child abuse. Conversely, increased opportunities and less burdensome role requirements for women should lower the probability of abuse. Applying this view to female participation in the paid labor force produces two competing hypotheses. First, if women voluntarily participate in the labor force and have help caring for children and the household, their participation may indicate they are taking advantage of available options, which would presumably lower levels of stress and child abuse. However, if mothers move into the labor force out of economic necessity and receive little childcare or household support, then employment may increase stress in the household as well as child abuse. Hence, it can be argued either that female labor-force participation decreases child abuse by providing greater alternatives to mothers, or that it increases abuse by increasing tensions in family relationships.

The first possibility would likely apply to women working in occupations that provide avenues for personal gratification and income, including the ability to purchase quality childcare. The second possibility seems more plausible for overall female labor-force participation in light of the fact that women are generally concentrated in low-paying/low-prestige jobs, and that many working mothers—especially those who are the sole economic providers for their families—are in the labor force out of economic necessity. There is some indirect research supporting this second argument. In a study of child abuse reported to authorities in 58 New York counties, Garbarino (1976) showed that of 12 variables analyzed, the percentage of women in the labor force was the most highly correlated with child abuse. He concluded that counties with higher rates of female labor-force participation included large

numbers of "economically depressed mothers, often alone in the role of parent, attempting to cope in isolation without adequate facilities and resources for their children" (p. 183). Similarly, in a U.S. study of married couples in which the husband was currently employed, Kessler and McRae (1982, p. 220) found that female participation in the work force reduced psychological distress only for those mothers whose husbands shared childcare duties. The authors concluded (p. 223) that "the benefits of employment are tied to conditions in the home (especially childcare responsibilities) and in the workplace" (see also Gelles and Hargreaves 1981; Bronfenbrenner and Crouter 1982; Belsky 1984).

Such findings suggest that social programs that lessen the impact of economic stress on the family should reduce the incidence of abuse. Thus, Zigler and Muenchow (1983) and Kamerman (1980) argue that part of the reason for high rates of child abuse in the United States is the lack of economic and social support for families with young children. Similarly, Vesterdal (1977) argues that an important reason for low rates of child abuse in Denmark is the widespread availability of social and economic support for mothers.

Social Disorganization

Social-disorganization theory first gained popularity in the United States in the early part of the twentieth century (e.g., Shaw and McKay 1932; Sellin 1938; Shaw, McKay, and McDonald 1938) and can be traced in part to Durkheim's ([1893] 1947) assessment of the transition to a modern industrial society. According to this perspective, deviance and social pathology result from the disruption of institutionalized patterns of social behavior. Such disequilibrium may be linked to the *level* of development, although the more common view is to see it associated with *rapid* social change that leaves insufficient time for new patterns of social organization to emerge. The disruption of established systems of role allocation and the emergence of new roles not yet fully institutionalized and integrated into society make normative guidelines ambiguous and may disrupt traditional support mechanisms.

This perspective suggests that child homicide may be more likely in societies undergo-

ing rapid social change, particularly when traditional patterns have been disrupted through economic growth, urbanization, and the shift from agricultural to industrial and service economies. The social-disorganization perspective has been most frequently applied to lesser-developed countries experiencing rapid social change. Although we could find no explicit cross-national tests of this assertion, several sources suggest such a relationship based on reviews (e.g., Gelles and Pedrick-Cornell 1983; Gelles 1987; Korbin 1987) or case studies of lesser-developed societies, including Nigeria (Jinadu 1980; Oyemade 1980), the Zulu tribe (Loening 1981), and India (Bhattacharyya 1979).

Culture of Violence

Early theorists like Shaw and McKay (1932) assumed that the criminal values and traditions that develop in socially disorganized areas are self-perpetuating. This idea finds more recent expression in subculture-of-violence theory. For example, Wolfgang and Ferracuti (1967) argue that violence, manifested in such criminal acts as homicide and assault, is part of a more general subculture of values and norms that legitimate the use of violence. Although such reasoning is used by Wolfgang and Ferracuti to suggest that certain societal subcultures may be prone to violence, similar reasoning is used to explain high rates of child abuse within entire nations (e.g., Gil 1975, p. 350; Gil 1970; Straus 1980; Gelles 1987).

We were not able to identify any comparative studies of child abuse that explicitly test these claims. However, if child homicide is correlated with a more general acceptance of violence and aggression, societies with higher levels of child homicide should also have higher levels of adult homicide. This hypothesis was tested by Christoffel et al. (1981), who found no correlation between general homicide rates and rates for children less than one year old, for either 24 developed or 23 developing nations. For children between the ages of one and four, they found a correlation between homicide rates for the developed but not the developing nations. Although these results do not support the culture-of-violence argument, they were limited to bivariate analyses.

Social Isolation and Surveillance

The idea that social isolation of families

increases child abuse has become common. Young (1964) concluded that 80–95 percent of abusing families in his U.S. study lacked continuing relationships with nonfamily members. Among poor families in three ethnic groups, Giovannoni and Billingsley (1970, pp. 196–204) found that families that did not participate in cooperative neighborhood activities (especially shared homemaking and child care) were more likely to be abusive or neglectful. Cross-cultural studies of parenting (Minturn and Lambert 1964; Rohner 1975) suggest that mothers who are isolated in childcare tasks are harsher with their children.

Two related explanations for an association between isolation and abuse appear in the literature. The most common is that isolation increases social and economic stress between caregivers and their children and increases the likelihood and seriousness of abuse. Garbarino (1977, p. 569) argues that greater isolation reduces alternative childcare resources and emergency economic support (see also Freeman 1979, p. 28). A second explanation is that isolation reduces the effectiveness of surveillance and, hence, allows increased abuse. For example, Garbarino (1977, p. 569) argues that greater social contact provides "personalized monitoring of the family." This explanation suggests that child abuse may increase as contact with family members, peers, and representatives of religious, civic, and governmental groups declines.[1] Although researchers have rarely used these ideas to explain cross-national variation in child abuse, it is reasonable to expect that societies in which child caregivers are more isolated from the influence (and/or surveillance) of family, friends, and social control agents should have higher levels of child abuse.

_____

[1] This argument seems compatible with recent criminology research on "situational" crime perspectives (e.g., Clarke, 1980; Felson and Cohen, 1980). Cohen and Felson (1979, p. 589) argue that violent crimes against the person require three basic elements: motivated offenders, suitable targets, and the absence of capable guardians. In general, crime can only take place when these three elements co-occur. Child homicide is unusual in that it is most often committed by one of the child's caregivers. Thus, two of the three elements, the offender and the target, are frequently in close contact with each other, and the issue of capable guardians and social isolation becomes particularly important.

DATA AND METHODS

Our measures of child homicide are (1) known homicides per 100,000 live births for children less than one year old; and (2) known homicides per 100,000 population for children one to four years old (World Health Organization 1976; 1977). Only homicides purposely inflicted resulting in legal intervention are included.[2] To examine the economic-stress view, we measure the effects of inequality (H1, hypothesis 1) by using data on the percentage of national income going to the poorest 20 percent of families around 1965, as well as the ratio of income of the top 20 percent to the bottom 40 percent (Bollen and Jackman 1985b).[3] The unemployment rate for 1970 (H2) is from the World Bank (1980). We measure the status of women in society (H3) by the ratio of females to males in professional employment in 1960 (ILO 1972)[4] and by the percentage of tertiary students who were female in 1970 (UNESCO 1975). Both measures reflect the level of women's access to status-granting and decision-making institutions. Women's overall labor-force participation (H4) is measured by female share of the labor force in 1965 (ILO 1977). Societal attempts to limit the impact of economic deprivation (H5) is measured by government revenue as a percentage of gross domestic product (GDP) in 1973 (World Bank 1976), social security expenditures as a percentage of gross national product (GNP) in 1974 (ILO 1981), and social-security family expenditures as a percentage of GNP in 1974 (ILO 1981).[5]

To examine the social-disorganization perspective, we measure the *level* of development (H6) by per capita GNP in 1970 (IBRD 1976), the *speed* of social change (H7) by average urban growth rates between 1960 and 1970 (World Bank 1982), and the ratio of per capita GNP in 1975 to per capita GNP in 1960 (IBRD 1976). After Christoffel et al. (1981), we examine the culture-of-violence perspective by using the total homicide rate for 1974 as an indicator of the degree to which violence is culturally accepted (H8).

We examine the surveillance view by using the state revenue and expenditure variables to indicate surveillance at the governmental level (H9), physicians per 100,000 population (WHO 1977) as an indicator of community surveillance (H10), and the average number of persons per household (United Nations 1974) as a measure of surveillance by close acquaintances and the family (H11).

We use regression analysis with intercept and slope dummy variables (Hanushek and Jackson 1977, pp. 106–8) to examine the effects of independent variables on child homicide in more-developed and lesser-developed countries (MDCs and LDCs).[6]

---

[2] Death rates were for 1974 for all but five countries that appear in analyses (El Salvador, Nicaragua, and Paraguay were for 1973; Colombia and Peru were for 1972). The correlation between the two measures for analyses with the most cases is .58 (N=40).

[3] Following Bollen and Jackman (1985b), equations examining income distribution contained a control variable indicating if an observation for the income measure was based on individuals or families.

[4] Data availability caused us to use 1960 data instead of the preferable 1970 or 1965 data. We assume the 1960 measure is substantially correlated with later measures.

[5] Clearly, not all government revenues are used to reduce the effects of economic hardship. However, a substantial percentage of government expenditures is targeted for welfare, at least in the

more-developed countries where the correlation between government revenue as a percentage of gross domestic product and social-security expenditures as a percentage of gross national product is .72 (N=18). The government revenue measure has the advantage of being available for a substantial number of both more-developed and lesser-developed countries. Social-security expenditures is clearly a general measure of welfare efforts, although it is available only for more-developed countries. Family expenditures is the most direct measure of assistance to modify the effects of inequality and economic stress on families with children, yet it is available for an even smaller number of developed countries than social-security expenditures. All three measures were not lagged because we expected the effects of income supplements, health care, counseling. and other spending to be almost immediate.

[6] The intercept dummy is constructed by creating a variable in which MDCs have a value of 0 and LDCs a value of 1. The slope dummy is constructed by multiplying the value of each independent variable by the intercept dummy. The slope dummy indicates how the effect of an independent variable on child homicide in LDCs differs from its effect in MDCs. Parameter estimates for independent variables in LDCs are. obtained by adding the parameter estimates of the

Table 1. Cross-sectional Regressions Examining the Effects of GNP and Alternative Independent Variables on Known Child Homicides for Children Less than One Year Old, and One to Four Years Old

EQ[#]	Dependent Variables[a]	Independent Variables[b]													R²	N
		Female Tertiary Enroll. Ratio, 1970			LDC Female Tert Enroll.			GNP/Capita 1970			LDC GNP/Capita					
		b	s.e.	t	b	s.e.	t	b	s.e.	t	b	s.e.	t			
1.	H <1	−.15**	.08	−1.79	.12	.10	1.14	.25	1.37	.18	−.15	1.63	−.09	.18	39
2.	H 1–4	−.00	.02	−.12	−.01	.03	−.52	.59*	.38	1.56	−.64*	.45	−1.42	.17	39
		Ratio F:M Professional, 1960			LDC F:M Professional											
3.	H <1	−7.83**	3.49	−2.24	9.09***	3.95	2.30	−.27	1.45	−.19	1.11	1.74	.64	.22	36
4.	H 1–4	−1.08	.98	−1.11	.55	1.11	.50	.56*	.41	1.37	−.54	.49	−1.12	.22	36
		Female Share Labor Force, 1965			LDC Female Share L.F.											
5.	H <1	.37***	.09	4.03	−.38***	.11	−3.52	−1.08*	.66	−1.64	1.30***	.40	3.25	.37	40
6.	H 1–4	.07***	.03	2.53	−.06**	.03	−1.96	−.10	.20	−.51	.23**	.12	1.88	.23	40
		Gov. Rev./GDP 1973			LDC Gov. Rev./GDP											
7.	H <1	−.26***	.07	−3.47	.41***	.11	3.70	3.75***	1.47	2.55	−4.15***	1.68	−2.47	.37	39
8.	H 1–4	−.08***	.02	−4.29	.06**	.03	2.20	1.63***	.34	4.29	−1.54***	.43	−3.56	.46	39
		Soc. Sec. Spend/ GNP 1974														
9.	H <1[c]	−.27**	.12	−2.27				1.26	1.85	.68				.24	19
10.	H 1–4[c]	−.09***	.03	−3.13				.87**	.46	1.89				.41	19
		Soc. Sec. Fam. Allow./GNP, 1974														
11.	H <1[c]	−1.66**	.82	−2.02				−.78	2.05	−.38				.24	16
12.	H 1–4[c]	−.44**	.19	−2.26				.13	.49	.27				.30	16
		GNP Growth 1960–75			LDC GNP Growth											
13.	H <1[c]	4.12*	2.99	1.38	−9.25*	3.82	−2.43	1.37	1.26	1.08	−.56	1.52	−.37	.24	40
14.	H 1–4[c]	.81	.85	.96	−.05	1.08	−.04	.67**	.36	1.87	−.74**	.43	−1.72	.19	40

[a] H = Known homicide rate   <1 = Less than one year old   1–4 = One to four years old   EQ[#] = Equation number
[b] b = Unstandardized regression coefficients   * = $p < .10$, one-tailed test
s.e. = Standard errors   ** = $p < .05$
t = t-ratios   *** = $p < .01$
   Main effects dummy variable included in all equations except 9 to 12. Results are not presented to conserve space.
[c] Developed countries only. No dummy variables included.

This method allows us to assess differences in the effects of independent variables in sample subpopulations. It is appropriate for the present investigation because (1) the social disorganization argument may apply more to LDCs than MDCs; (2) the greater resources of MDCs may make state welfare intervention more important in MDCs than LDCs; and (3) data-collection capacities in MDCs are greater than LDCs, making data from MDCs more reliable. We use United Nations classification procedures (U.N. 1973) to define a country as an LDC or MDC.

Because most theoretical approaches assume the level of economic development has a substantial influence on the character of society, we include log per capita GNP as a control for level of development in all equations. Our initial analyses include the intercept dummy, log per capita GNP and its slope dummy, and an alternate independent variable with its corresponding slope dummy.

RESULTS AND DISCUSSION

In Table 1 we report results of analyses in which the independent variables of interest were significant below the .10 level.[7] In

slope dummy to the parameter estimates for the corresponding variable in MDCs.

[7] Because child homicide for one to four year olds is divided by population and not the number of one to four year olds, homicide rates for countries with younger populations will be some-

general, the results show substantial support
for the economic-stress perspective in MDCs
and little support for social-disorganization,
culture-of-violence, or isolation perspectives
in either MDCs or LDCs. However, support
for the economic-stress perspective is selec-
tive, centering on the status of women and
societal efforts to reduce economic stress, not
income inequality and unemployment.

Perhaps the most striking finding in Table 1
is that greater female labor-force participation
results in higher homicide rates for both
children less than one (eq. 5) and children one
to four (eq. 6). As suggested in the theoretical
review, we interpret this finding as indicating
higher levels of role strain and stress for
family members when women move into paid
employment because of economic necessity.
At the same time, the results provide
substantial (though not total) support for the
hypothesis that countries that provide women
with greater opportunities have lower rates of
child homicide. Thus, higher participation by
women in tertiary education (eq. 1) and a
higher ratio of women to men in professional
jobs (eq. 3) are both associated with lower
homicide rates among children less than one
year old. Although nonsignificant, the results
are in the same direction for children one to
four years old (eqs. 2 and 4).

Results for the government-revenue, social-
security, and family-allowance measures are
also consistent with the economic-stress
argument. All three measures show that
developed countries providing increased pub-
lic assistance have lower child-homicide rates
(eqs. 7–12). The fact that measures of state
expenditures, but not economic inequality,
affect child-homicide rates may suggest that
state efforts to modify inequality are not
reducible to direct income measures. That is,
child-homicide rates appear to depend not on
income distributions per se, but rather on the
range of programs that provide aid to women,
children, and families, and thereby offer

important support, as well as a wider range of
choices, to parents.

Results for the social-disorganization per-
spective indicate that the level of develop-
ment may have a modest effect on child
homicide, while the rate of development has
no effect. Table 1 (column 3) shows that
countries with greater per capita GNP gener-
ally have higher rates of homicide. Per capita
GNP is positive in six of seven equations for
the one-to-four-year-old category and is
significant in five of these six equations. A
positive coefficient also appears in four of
seven equations for children less than one
year old, although it is significant in only one
equation. In contrast, urban growth is nonsig-
nificant, while GNP growth is positive yet
significant only for children less than one year
old (eq. 13).

Culture-of-violence and isolation perspec-
tives also receive little support. As Christoffel
et al. (1981) found, the overall homicide rate
in a country has no effect on the child-
homicide rate. Assuming that government
revenue and expenditure measures partly
capture the degree to which the state serves as
a surveillance agency, the results are consis-
tent with the hypothesis that state surveillance
reduces the child-homicide rate. However,
neither of the measures of community and
family isolation had an effect on child-
homicide rates.

The results for the slope variables corre-
sponding to each of the alternative indepen-
dent variables (column 2) and per capita GNP
(column 4) suggest that none of the indepen-
dent variables are good predictors of child
homicide in LDCs. Because no variable can
explain variation in child homicide among
LDCs, we exclude LDCs from the subsequent
discussion and analysis.[8]

Overall, results in Table 1 show that high
levels of female labor-force participation,
coupled with low levels of female status and
welfare spending, result in high child-
homicide rates. In contrast, inequality and

what inflated. To control for the size of the
one-to-four-year-old population, all equations in
Table 1 were estimated controlling for the role of
population growth, measured by population size in
1975 divided by population in 1960 (World Bank,
1976). Population growth was always positive and
often significant. However, inclusion of population
growth did not alter the effects of any of the other
variables on child homicide.

[8] One exception was the effect of the slope
dummy for GNP growth on homicides of children
less than one year old. Results indicate a modest
negative effect ($p < .10$, two-tailed test). While
such an unanticipated result could be a statistical
anomaly, it may also indicate that rapid growth
helps reduce economic stress by increasing eco-
nomic prosperity, thereby lowering child-homicide
rates.

Table 2. Cross-sectional Regressions Examining the Effects of GNP, Female Share of Labor Force, Female Tertiary Enrollment Ratio or Female Professional Enrollment Ratio, and One of Three Measures of State Spending on Known Child Homicides for Children Less than One Year Old, and One to Four Years Old. Analysis on Developed Countries Only

EQ[#]	Dependent Variables[a]	Independent Variables[b]													r^2	N
		GNP/Capita 1970			Female Share of Labor Force, 1965			Female Tertiary Enroll. Ratio, 1970			Gov. Rev./GDP 1973				
		b	s.e.	t	b	s.e.	t	b	s.e.	t	b	s.e.	t		
1.	H <1	.41	1.16	.35	.33***	.07	4.92	−.15***	.05	−2.86	−.16***	.05	−2.86	.82	18
2.	H 1–4	1.33***	.45	2.93	.04*	.03	1.48	−.00	.02	−.06	−.07***	.02	−3.56	.67	18
											Soc. Sec. Fam. Allow./GNP, 1974				
3.	H <1	−2.12*	1.37	−1.54	.40***	.10	3.87	−.20***	.07	−2.90	−.03	.77	−.05	.75	15
4.	H 1–4	−.01	.51	−.03	.05	.04	1.27	−.00	.03	−.18	−.37	.29	−1.27	.45	15
											Soc. Sec. Spend/GNP 1974				
5.	H <1	−.73	1.17	−.63	.36***	.07	4.92	−.17***	.05	−3.18	−.16**	.08	−2.16	.77	18
6.	H 1–4	.65*	.48	1.35	.05*	.03	1.70	−.00	.02	−.21	−.08***	.03	−2.52	.52	18
								Ratio F:M Professional, 1960							
7.	H <1	−1.26	1.30	−.97	.37***	.07	4.92	−6.84***	2.25	−3.04	−.18**	.08	−2.32	.79	17
8.	H 1–4	.87**	.50	1.73	.05**	.03	1.86	−.71	.87	−.81	−.09***	.03	−3.00	.60	17

[a] H = Known homicide rate   <1 = Less than one year old   1–4 = One to four years old   EQ = Equation number
[b] b = Unstandardized regression coefficients   * = p<.10, one-tailed test
s.e. = Standard errors   ** = p<.05.
t = t-ratios.   *** = p<.01.

unemployment, more commonly associated with child abuse in the literature, have no impact on child homicide. Similarly, variables drawn from other theoretical perspectives on child abuse—social disorganization, culture of violence and social isolation—have little or no impact on child homicide. The only variable showing significant effects that is not included under the economic-stress perspective is per capita GNP. We turn next to a model that includes only the variables found to be significant in predicting the child homicide rate in MDCs.

Table 2 presents four pairs of equations, each of which estimates the effect of per capita GNP, female share of labor force, women's status (either the female tertiary enrollment ratio or the ratio of female to male professionals), and government welfare efforts (government revenue, social-security spending, and social-security family allowance) on homicides of children less than one year old and one to four years old.

Results in Table 2 clearly show the importance of female labor-force participation, women's status, and government revenue and expenditures for predicting child homicide. Overall, women's labor-force par-

ticipation continues to have a positive effect on the child-homicide rate for both children less than one year old and one to four years old, a result that is statistically significant in seven of eight equations. In contrast, when women have higher social status (as measured by either tertiary enrollment ratio or ratio of female professionals), homicide rates for children less than one are significantly lower. Likewise, child-homicide rates are consistently lower in countries with greater levels of welfare spending, although results are not statistically significant when the family-allowance measure is used (see Appendix A). Results in Table 2 remain consistent even when outlying cases are deleted (see Appendix B).

In sum, analyses point to the utility of focusing on women's labor-force participation and status and on measures of state welfare efforts. They are consistent with the view that among developed countries limited opportunities for women, as well as economic stress accompanying high levels of female
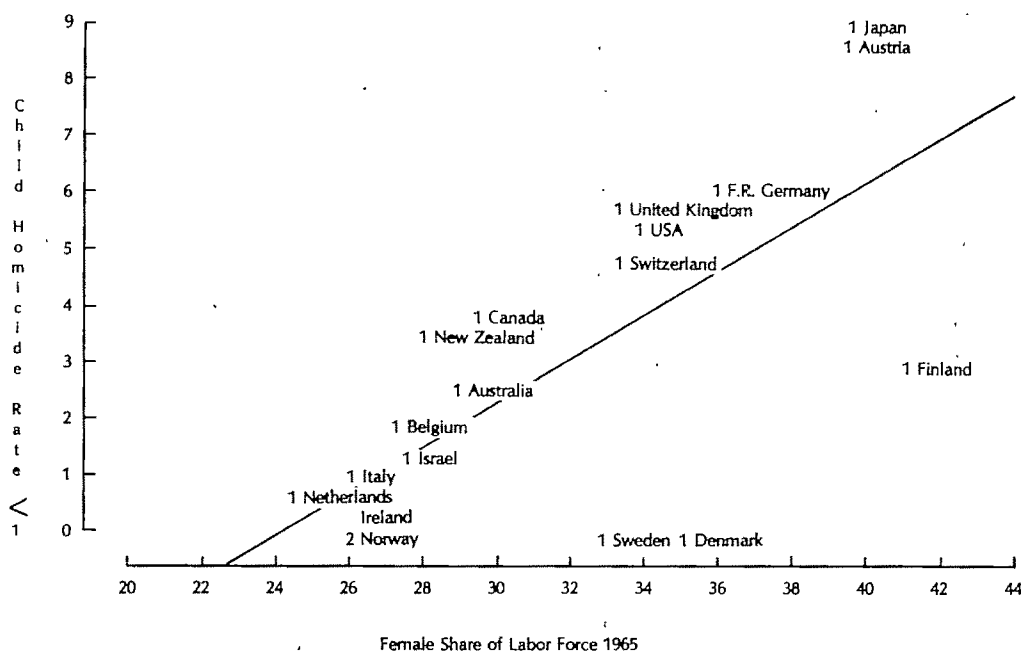
Residual values from equation 5 in Table 2 in standard deviation units.

Israel	-.14	F.R. Germany	1.00	United Kingdom	1.44	Ireland	-2.30
Japan	-.22	Netherlands	1.14	Finland	-1.47	Canda	2.58
New Zealand	.32	Italy.	1.32	Norway	-1.77	Denmark	-3.61
Sweden	-.92	Belgium	1.40	Austria	-1.92		
Australia	.93	Switzerland	-1.44	USA	2.23		

Fig. 1. Plot of Female Share of the Labor Force 1965 on Child Homicide Rate for Children Less Than One Year Old.

labor-force participation and low levels of state support, contribute to high child-homicide rates.[9]

Perhaps the most consistent finding in our analysis is the effect of female labor-force participation on child-homicide rates. Given the relatively small number of cases in the analysis, a bivariate plot of these variables may help clarify their relationship.[10] In Figure 1, we plot female labor-force participation on the homicide rate for children less than one, using the 18 cases analyzed in the equations in Table 2. To help interpret the plot, we include a note on residual values for individual countries from predicted scores of the full model, estimated from equation 5 in Table 2. The residuals are rank ordered by distance in standard deviation units from the regression line.

Figure 1 shows a remarkably clear positive relation between female labor-force participation and child homicide. The six countries with the highest homicide rates all have relatively high female labor-force participa-

---

[9] Such an interpretation receives additional support from analyses where measures of the percentage of married and single women in the labor force (United Nations, 1969) are placed in the same equation. In analyses similar to those in Table 1 ($N = 24$), the proportion of females working was positive and significant at less than the .01 level on both homicide measures, while the proportion of single females working was positive and insignificant (t never greater than .62) on both measures.

[10] Of course, bivariate plots will be misleading to the extent that relationships between independent and dependent variables change substantially from bivariate to multivariate models. To determine the viability of examining a bivariate plot of

female labor-force participation and child homicide, we estimated a series of bivariate equations that examined the effect of each of the variables in Table 2 on the two measures of child homicide, using the same cases found in the equations in Table 2. The results (available on request) showed remarkable continuity with those obtained from the multivariate model, indicating that bivariate plots may be usefully interpreted.

tion rates, while four of the six countries with the lowest homicide rates have relatively low female labor-force participation rates. Additionally, the modest residual values indicate that the model fits most countries fairly well. It is instructive to consider why Sweden and Finland, with homicide rates substantially less than their female labor-force participation rates would imply, have such modest residual values from the full equation ($-.92$ and -1.47 respectively). The answer is that their values on social-security expenditures and female tertiary enrollment are relatively high. Thus Sweden, with a low homicide rate, has nearly the highest value on each of these two measures, suggesting that a commitment to family support helps mitigate individual and family stress accompanying child care responsibilities. Similarly, Finland, with a midrange homicide rate, has the highest enrollment ratio, while its value on social security is at the center of the distribution. Although Denmark exhibits the largest residual value from the full equation, this residual would likely have been greater if it did not have high values on female tertiary enrollments and social-security expenditures. Denmark's greater residual than Sweden, which has the same homicide rate, is due to Denmark's lower values for female tertiary enrollment and social-security expenditures.

Few people will be surprised by such results for the Scandinavian countries, a region widely recognized as having among the most comprehensive social-welfare measures in the world. Sweden provides working mothers not only with childbirth leave, but also with time off to care for sick children. Moreover, Swedish law allows either parent a six-hour workday, with income supplements until the child is eight years old (Kamerman 1980, p. 131; Zigler and Muenchow 1983). Similar policies are well established in Finland and Denmark. Lindgren (1978, pp. 286–87) reports that maternity leave in Finland is currently 174 days, during which time the mother is entitled to a maternity allowance of 30 to 50 percent of her average daily pay. If the mother is the principal or sole family earner, she receives an additional allowance. Even mothers who are not economically active are eligible for payments. Likewise, Danish family policy includes a generous family allowance for all children, a national system of subsidized daycare institutions, and free medical examinations for

children (Wagner and Wagner 1976; Vedel-Petersen 1978).

The other cases in Figure 1 fall fairly close to the regression line for female employment and have modest residual values from the full equation. New Zealand shows the smallest residual value from the estimate of the full equation. Its homicide rate is not only close to the regression line for female labor-force participation, but is even closer to the unreported regression lines for female tertiary enrollments and social-security expenditures. Like patterns exist for Israel, the Netherlands, Australia, Italy, Switzerland, and Belgium. The remaining cases fit the model and/or plots less well, yet do not substantially contradict the empirical portrait presented earlier. Still, much could be learned by thoughtful comparative consideration of individual countries.

For example, although Japan fits the model quite well, its high homicide rate may be influenced by factors other than its relatively high levels of female labor-force participation, low levels of state welfare expenditures, and low tertiary enrollments. Specifically, it has been argued that, while the rate of child abuse in Japan is relatively low, the incidence of homicide as a form of abuse is fairly high. Relatedly, it has been suggested that *oyako shinju*—in which the parent kills the child and then commits suicide—may be a uniquely Japanese phenomenon (Wagatsuma 1981, p. 121). Placing such case study material in a comparative framework and combining it with cross-national studies could help sharpen our view of the social forces influencing child abuse and homicide.

While this investigation provides a plausible explanation for cross-national variation in child-homicide rates, various limitations point to avenues for future research. These limitations include (1) the cross-sectional research design and its limited temporal span; (2) the operationalization of major concepts; (3) the focus on a single dramatic form of child abuse; and (4) the possibility of less-reliable data for LDCs. Time series and historical studies could help clarify causal processes and changes in the dynamics of child abuse through historical epochs, especially if such inquiries are comparative. This could be accomplished either through the analysis of multiple cases or the judicious use of information from other cases when examining individual countries. Although we have attempted to use appropriate indicators of major

concepts, alternative ones are clearly possible and may provide a valuable check on the present findings. Additionally, while cross-national studies of forms of abuse other than homicide may be difficult, comparative case studies could examine a major assumption of the present inquiry: that child homicide is related to the level of other forms of abuse in society. Future inquiries might find that, while child homicide is a good indicator of violence toward children, it performs poorly as an indicator of other forms of abuse. Finally, future research should look more closely at the absence of effects for variables in LDCs. These results may be a consequence of less-reliable data, or perhaps the dynamics of child homicide in LDCs are so diverse as to make generalizations difficult. Future work on LDCs should be sensitive to this issue and may profitably attempt to work somewhat outside the dominant social-disorganization perspective.

Beyond these largely methodological concerns is a need for reformulating macrosociological explanations of abuse, perhaps by combining and refining existing perspectives. For example, part of a reformulation consistent with our results would point to the manner in which an increasingly rationalized world economy, accompanied by an ideology of equality (Bendix 1978; Meyer 1980) and the notion that progress occurs through the mobilization of individuals (Ramirez and Boli 1987; Fiala and Gordon Lanford 1987), has extended citizenship status to women by including them in the political apparatus and the paid labor force (Ramirez and Weiss 1979; Ramirez and Boli 1982). Yet such changes have not been uniformly accompanied by corresponding changes in the state, economy, and role structure. Thus, role strain and ambiguity develop, which, in turn, increase individual and family stress and the likelihood of child abuse and homicide. This argument incorporates a focus on female status and economic stress within a broader framework that also includes elements of the social-disorganization and isolation perspectives. Other such reformulations may be possible, yet the underlying task remains the same: to develop a coherent macrosociological explanation of the social sources of abuse.

CONCLUSION

This investigation used cross-national data to overcome some of the conceptual and data limitations that have hampered previous comparative studies of child abuse. Analysis of hypotheses emerging from four theoretical perspectives indicated that a focus on women's status and economic stress was most useful in explaining national variation in child-homicide rates. Our results suggest that the economic stress accompanying female labor-force participation contributes to increased national child-homicide rates, while more equal participation of women in society and greater state efforts to modify the impact of economic stress on the family may reduce homicide rates. Hypotheses from social-disorganization, culture-of-violence, and social-isolation perspectives received either modest or no support. No perspective provided hypotheses able to explain variation in child-homicide rates in LDCs.

Although the proportion of mothers who work for pay has increased dramatically in recent decades, most countries have been slow to adjust. Many societies that have lowered obstacles to female participation in the paid labor force have simultaneously ignored or deemphasized the problems that naturally accompany this policy. Part of the problem may be solved as women achieve more equal participation in the paid labor force—the burdens of child care will be less stressful for mothers who can afford to buy adequate support. However, even if the quality of women's labor-force participation were identical to men's, serious difficulties would remain for single and poor parents. Given that women are clearly in the paid labor force to stay, policies that continue to ignore the problems faced by working parents are likely to have serious consequences for the world's children.

APPENDIX A: A Note on the Family-Allowance Measure

The nonsignificant results for the family-allowance measure in Table 2 are potentially important because we expect family allowance to be the most direct measure of state efforts to reduce the negative effects of economic hardship on families. We explored several possible explanations for the result. First, it could be a consequence of the low values and modest variation in the family-allowance measure (mean = 1.16 percent, SD = .77 percent), making errors and small variations in reporting procedures potentially problematic, especially in an analysis with only 10 degrees of

freedom. Second, compared to the other measures of state spending, data on the family-allowance measure were missing for an additional three countries. To evaluate the impact of these missing cases, we estimated values on the family-allowance measure for the three cases by first noting the rank-order position of the cases in the social-security data and then giving each case a family-allowance value midway between the family-allowance value of the case above and below it. Reanalysis of the equations in Table 2 using the estimated data produced a nonsignificant negative parameter estimate for children less than one $(-.50; t = -.79)$ and a significant negative estimate for children one to four years old $(-.52; t = 2.24)$. Third, Austria is a major outlier on the family-allowance measure, having high rates of both family allowance and child homicide. Analyses deleting Austria produced nonsignificant parameter estimates of $-.61$ $(t = -.56)$, and $-.23$ $(t = -.60)$ for children less than one and one to four respectively. Finally, it is possible that results for the family-allowance measure may be due to the fact that both family allowances and female tertiary enrollments are measuring a common variable: general social support for sexual equality and family well-being. The greater reliability of the tertiary-enrollment measure, and its greater variation, may cause it to be the better indicator of the common variable. Some support for this idea is found in the fact that the only drop in the family-allowance coefficient between Tables 1 and 2 occurs for children less than one—the same category in which tertiary enrollments show significant effects. When we estimated the equations without female tertiary enrollments, the family allowance estimate increased to $-.76$, although it remained nonsignificant $(t = -.98)$. While we are unable to draw firm conclusions about the family-allowance measure, it is important to note that in all analyses the family allowance measure was consistently associated with lower child-homicide levels.

APPENDIX B: Residual Analyses

To determine whether outlying observations are responsible for results, we identified outlying cases by examining residual plots (Morris and Rolph 1981) and partial regression plots (Bollen and Jackman 1985a). Observations two standard deviation units above or below predicted values were considered outliers, while those 1.75 standard deviations from predicted values were considered near outliers. These procedures identified the following outliers (O) and near outliers (N.O.: Equation 1: Denmark (O), Austria (N.O.), Finland (N.O.); Eq. 2: United Kingdom (N.O.), Australia (N.O.); Eq. 3: Austria (O), Belgium (O), Canada (O); Denmark (N.O.); Eq. 4: Austria (O), Japan (O), Canada (O); Eq. 5: Austria (O), Denmark

(O), Ireland (N.O.); Eq. 6: Austria (O), Israel (N.O.); Eq. 7: Austria (O), Canada (O), Sweden (O), Finland (O), Israel (N.O.), Italy (N.O.), Denmark (N.O.); Eq. 8: Israel (O), Finland (O), Norway (N.O.), Italy (N.O.), Canada (N.O.), Denmark (N.O.).

The equations in Table 2 were reestimated several times, eliminating one of the outlying or near outlying observations for each equation each time. After this, all the outliers and near outliers for equations 1 to 6 were deleted and the equation reestimated. For equations 7 and 8, outliers were not combined because of the small number of cases. These procedures produced results very similar to those in Table 2.

REFERENCES

Belsky, Jay. 1984. "The Determinants of Parenting: A Process Model." *Child Development* 55:83–96.

Bendix, Reinhard. 1978. *Kings or People.* Berkeley: University of California Press.

Bhattacharyya, A.K. 1979. "Child Abuse in India and the Nutritionally Battered Child." *Child Abuse and Neglect* 3:607–14.

Bollen, Kenneth and Robert Jackman. 1985a. "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases." *Sociological Methods and Research* 13:510–42.

———. 1985b. "Political Democracy and the Size Distribution of Income." *American Sociological Review* 50:438–57.

Breines, Wini and Linda Gordon. 1983. "The New Scholarship on Violence." *Signs* 8:490–531.

Bronfenbrenner, Urie and A.C. Crouter. 1982. "Work and Family Through Time and Place." Pp. 39–83 in *Families that Work,* edited by Sheila Kamerman and Cheryl Hayes. Washington, DC: National Academy Press.

Burgdorf, K. 1980. *Recognition and Reporting of Child Maltreatment.* Rockville, MD: Westat, Inc.

Christoffel, Katherine K. and Kiang Liu. 1983. "Homicide Death Rates in Childhood in 23 Developed Countries: U.S. Rates Atypically High." *Child Abuse and Neglect* 7:339–45.

Christoffel, Katherine K., Kiang Liu, and Jeremiah Stamler. 1981. "Epidemiology of Fatal Child Abuse: International Mortality Data." *Journal of Chronic Diseases* 34:57–64.

Clarke, R.V.G. 1980. " 'Situational' Crime Prevention: Theory and Practice." *British Journal of Criminology* 20:136–47.

Cohen, Lawrence E. and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44:588–608.

Durkheim, Emile. [1893] 1947. *The Division of Labor in Society.* Tr. by George Simpson. Glencoe, IL: Free Press.

Felson, Marcus and Lawrence E. Cohen. 1980. "Human Ecology and Crime: A Routine Activity Approach." *Human Ecology* 8:389–406.

Fiala, Robert and Audri Gordon Lanford. 1987. "Educational Ideology and the World Educational Revolution, 1950–1970." *Comparative Education Review* 31:315–32.

Freeman, M.D.A. 1979. *Violence in the Home.* Westmead, England: Saxon House.

Garbarino, James. 1976. "A Preliminary Study of Some Sociological Correlates of Child Abuse: the Impact of Socioeconomic Stress on Mothers." *Child Development* 47:178–85.

———. 1977. "The Price of Privacy: An Analysis of the Social Dimensions of Child Abuse." *Child Welfare* 56:565–75.

———. 1987. "The Consequences of Child Maltreatment: Biosocial and Ecological Issues." Pp. 299–315 in *Child Abuse and Neglect: Biosocial Dimensions,* edited by R. Gelles and J. Lancaster. New York: Aldine de Gruyter.

Gelles, Richard. 1974. *The Violent Home: A Study of Physical Aggression Between Husbands and Wives.* Beverly Hills: Sage.

———. 1975. "The Social Construction of Child Abuse." *American Journal of Orthopsychiatry* 43:611–21.

———. 1987. "What to Learn from Cross-Cultural and Historical Research on Child Abuse and Neglect: An Overview." Pp. 15–30 in *Child Abuse and Neglect: Biosocial Dimensions,* edited by R. Gelles and J. Lancaster. New York: Aldine de Gruyter.

Gelles, Richard J. and Eileen F. Hargreaves. 1981. "Maternal Employment and Violence Toward Children." *Journal of Family Issues* 2:509–30.

Gelles, Richard J. and Claire Pedrick-Cornell, eds. 1983. *International Perspectives on Family Violence.* Lexington, MA: Lexington Books.

Gelles, Richard J. and Murray A. Straus. 1979. "Determinants of Violence in the Family: Toward a Theoretical Integration." Pp. 549–81 in *Contemporary Theories About the Family,* vol. 1, edited by W. Burr, R. Hill, I. Nye, and I. Reiss. New York: Free Press.

Gil, David. 1970. *Violence Against Children: Physical Child Abuse in the United States.* Cambridge: Harvard University Press.

———. 1975. "Unraveling Child Abuse." *American Journal of Orthopsychiatry* 45:346–356.

Giovannoni, Jeanne M. and A. Billingsley. 1970. "Child Neglect Among the Poor: A Study of Parental Adequacy in Families of Three Ethnic Groups." *Child Welfare* 49:31–40.

Goode, William J. 1971. "Force and Violence in the Family." *Journal of Marriage and the Family* 33:624–36.

Hansmann, Henry B. and John M. Quigley. 1982. "Population Heterogeneity and the Sociogenesis of Homicide." *Social Forces* 61:206–24.

Hanushek, Eric and John Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic Press.

International Bank for Reconstruction and Development. 1976. *1976 Atlas: 1973–1975 Base Period.* Washington, DC: Economic Analyses and Projections Department.

International Labour Office (ILO). 1972. *Yearbook of Labor Statistics.* Geneva: ILO.

———. 1977. *Labor Force Estimates and Projections.* Geneva: ILO.

———. 1981. *The Case of Social Security, 1972–1974.* Geneva: ILO.

Jinadu, M. 1980. "The Role of Neglect in the Etiology of Protein-Energy Malnutrition in Urban Communities of Nigeria." *Child Abuse and Neglect* 4:233–45.

Kamerman, Sheila B. 1980. *Parenting in an Unresponsive Society: Managing Work and Family Life.* New York: Free Press.

Kempe, C. Henry, Frederic N. Silverman, Brandt F. Steele, William Doegemueller, and Henry K. Silver. 1962. "The Battered Child Syndrome." *Journal of the American Medical Association* 181:107–12.

Kessler, Ronald C. and James A. McRae Jr. 1982. "The Effects of Wives' Employment in the Mental Health of Married Men and Women." *American Sociological Review* 47:216–27.

Korbin, Jill. 1981. *Child Abuse and Neglect: Cross-Cultural Perspectives.* Los Angeles: University of California Press.

———. 1987. "Child Maltreatment in Cross-Cultural Perspective: Vulnerable Children and Circumstances." Pp. 31–56 in *Child Abuse and Neglect: Biosocial Dimensions,* edited by R. Gelles and J. Lancaster. New York: Aldine de Gruyter.

LaFree, Gary and Edward Kick. 1986. "Cross-National Effects of Developmental, Distributional, and Demographic Variables on Crime: A Review and Analysis." *International Annals of Criminology* 24:213–35.

Light, Richard J. 1974. "Abused and Neglected Children in America: A Study of Alternative Policies." *Harvard Educational Review* 43:556–98.

Lindgren, Jarl. 1978. "Family Policy as a Field: Finland." Pp. 270–94 in *Family Policy: Government and Families in Fourteen Countries,* edited by S. Kamerman and A. Kahn. New York: Columbia University Press.

Loening, W. 1981. "Child Abuse Among the Zulus: A People in Transition." *Child Abuse and Neglect* 5:3–7.

Maney, A.C. and Benjamin Kedem. 1982. "A Binary Time-Series Analysis of Domestic Child Homicide." *Evaluation Review* 6:393–402.

Messner, Steven F. 1982. "Societal Development, Social Equality, and Homicide: A Cross-National Test of a Durkheimian Model." *Social Forces* 61:225–40.

Meyer, John. 1980. "The World Polity and the Authority of the Nation State." Pp. 109–38 in *Studies of the Modern World System,* edited by A. Bergesen. New York: Academic Press.

Morris, Carl and John Rolph. 1981. *Introduction to Data Analysis and Statistical Inference.* Englewood Cliffs, NJ: Prentice Hall.

Minturn, L. and W. Lambert. 1964. *Mothers of Six Cultures: Antecedents of Childrearing.* New York: John Wiley.

Newberger, E.H., R.B. Reed, J.H. Daniel, J.N. Hyde, and M. Kotelchuck. 1977. "Pediatric Social Illness: Toward an Etiologic Classification." *Pediatrics* 60:178–85.

Oyemade, A. 1980. "Child Care Practices in Nigeria—An Urgent Plea for Social Workers." *Child Abuse and Neglect* 4:101–3.

Ramirez, Francisco and John Boli. 1982. "Global Patterns of Educational Institutionalization." Pp. 15–38 in *Comparative Education,* edited by P. Altbach, R. Arnove, and G. Kelley. New York: Macmillan.

———. 1987. "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization." *Sociology of Education* 60:2–17.

Ramirez, Francisco and Jane Weiss. 1979. "The Political Incorporation of Women." Pp. 238–52 in *National Development and the World System,* edited by J. Meyer and M. Hannan. Chicago: University of Chicago Press.

Rohner, Ronald. 1975. *They Love Me, They Love Me Not: A Worldwide Study of the Effects of Parental Acceptance and Rejection.* New York: Human Relations Area Files Press.

Sellin, Thorsten. 1938. *Culture, Conflict and Crime.* New York: Social Science Research Council.

Shaw, Clifford R. and Henry D. McKay. 1932. *Juvenile Delinquency in Urban Areas.* Chicago: University of Chicago Press.

Shaw, Clifford R., Henry D. McKay, and James F. McDonald. 1938. *Brothers in Crime.* Chicago: University of Chicago Press.

Sims, B.G. and J.M. Cameron. 1973. "Bite-Marks in the 'Battered Baby Syndrome.' " *Medicine, Science and the Law* 13:207–10.

Steele, Brandt F. 1977. "Child Abuse and Society." *Child Abuse and Neglect* 1:1–6.

Steinmetz, Suzanne K. 1978. "Violence Between Family Members." *Marriage and Family Review* 1:1–16.

Steinmetz, Suzanne K. and Murray A. Straus. 1974. *Violence in the Family.* New York: Harper and Row.

Straus, Murray. 1980. "A Sociological Perspective on the Prevention of Wife-Beating." Pp. 211–32 in *Social Causes of Husband-Wife Violence,* edited by M. Straus and G. Hotaling. Minneapolis: University of Minnesota Press.

Straus, Murray, Richard Gelles, and Susan Steinmetz. 1980. *Behind Closed Doors: Violence in the American Family.* Garden City, NY: Anchor Press, Doubleday.

UNESCO. 1975. *Statistical Yearbook.* Louvain, Belgium: UNESCO.

United Nations. 1969. *Demographic Yearbook, 1968.* New York: United Nations.

———. 1973. *Statistical Yearbook.* New York: United Nations.

———. 1974. *Compendium of Housing Statistics.* New York: United Nations.

Vedel-Petersen, Jacob. 1978. "Family Policy as a Field: Denmark." Pp. 295–327 in *Family Policy: Government and Families in Fourteen Countries.* New York: Columbia University Press.

Vesterdal, Jorgen. 1977. "Handling Child Abuse in Denmark." *Child Abuse and Neglect* 2: 193–98.

Wagatsuma, H. 1981. "Child Abandonment and Infanticide: A Japanese Case." Pp. 120–38 in *Child Abuse and Neglect: Cross Cultural Perspectives,* edited by J.E. Korbin. Berkeley: University of California Press.

Wagner, Marsden and Mary Wagner. 1976. *The Danish National Child-Care System.* Boulder. CO: Westview Press.

Williams, Gertrude. 1980. "Toward the Eradication of Child Abuse and Neglect at Home." Pp. 588–605 in *Traumatic Abuse and Neglect of Children at Home,* edited by Gertrude Williams and John Money. Baltimore: Johns Hopkins University Press.

Wolfgang, Marvin E. and Franco Ferracuti. 1967. *The Subculture of Violence.* London: Tavistock.

World Bank. 1976. *World Tables.* Baltimore: Johns Hopkins Press.

———. 1980. *World Tables, 1980.* Baltimore: Johns Hopkins Press.

———. 1982. *World Development Report.* New York: Oxford University Press.

World Health Organization. 1976. *World Health Statistics Annual.* Geneva: World Health Organization.

Young, L. 1964. *Wednesday's Children: A Study of Child Neglect and Abuse.* New York: McGraw-Hill.

Zigler, Edward and Susan Muenchow. 1983. "Infant Day Care and Infant-Care Leaves." *American Psychologist* 38:91–94.

RESEARCH NOTES

◆

SOCIAL CONTROL AND THE DIFFUSION OF MODERN TELECOMMUNICATIONS TECHNOLOGIES: A CROSS-NATIONAL STUDY*

BRADLEY JAY BUCHNER

University of South Carolina

Published data routinely show a notable disparity in the relative growth of telephone and television technologies between Marxist and non-Marxist industrial nations, with Marxist nations favoring growth of television systems and non-Marxist nations generally favoring telephone systems. A study of 8 Marxist and 23 non-Marxist industrial nations was conducted using simple descriptive techniques and regression analysis. Results indicated that regime practices were a much stronger predictor than relative economic development.

INTRODUCTION

Communications technologies have developed and proliferated at a furious pace in the 20th century. Yet the development and dispersion of various types of communications technologies have proceeded in different directions in different societies. In this paper I examine the differential growth and diffusion of telephone and television technologies among Eastern and Western European countries, as well as in other market-industrial nations.

Statistics (Statistical Abstract of the United States 1985)[1] show that Eastern European nations have, without exception, more televisions than telephones. Why should this be so? In the absence of relevant information, one could argue that this is simply a matter of

---

[1] The original idea for the research came from looking at data on communications found in the chapter on Comparative International Statistics in the *Statistical Abstract*. However, the data used in the study are taken from Taylor and Jodice (1983). For further information on the sample and the specific data manipulations used, write to the author at the address listed above.

consumer preference. Television is an attractive means of entertainment. I shall demonstrate that this explanation is not satisfactory and that the basic structure of regimes is a better one.

The different functions of telephone and television are of considerable theoretical importance. Television broadcasts emanate from one source and constitute a one-way system of communication. Although the telephone requires a central exchange, it permits the free flow of two-way communication and makes control by the central office a difficult and expensive affair. These technologies represent two basic types of communication, public and private.

My central hypothesis is that states that practice a high degree of central authority favor the diffusion of television over telephones. I hypothesize that Marxist East European governments favor the growth of television over telephones. More importantly, this favoritism cannot be satisfactorily explained by conventional factors such as GNP per capita, population density, energy consumption, or urbanization. On the other hand, nations favoring decentralized government with a "free" market either show no favoritism, or favor the growth of telephones over televisions because interpersonal communication is valued and useful to the operation of a market economy.

While other scholars (cf. Bower 1985; McLuhan 1964) focus on the *consequences* of diffusion, without regard to political settings, this research addresses the *causes* of differential diffusion of communications technologies

in different political settings. Also, while most researchers focus on how communications affects economic growth, I am interested in how political regimes employ modern telecommunications in the pursuit of societal goals. Previous research examines television and telephone systems separately, while I examine the systems as complementary means of communication that have different impacts on the social organization of societies.

The sample for this study is 31 nations classified by the World Bank as "market-industrial" and "non-market industrial." The nations are grouped into "Marxist" and "non-Marxist" categories for this analysis. While these categories do not allow for the variability within the two groups of nations, that is not my central concern.[2]

THE PROBLEM AND THE LITERATURE

The two technologies offer an important contrast. Television is an extremely attractive tool in exposing a target population to a product, political viewpoint, or ideology. The telephone, on the other hand, has little educational capability and is more difficult to control. Given limited resources, regimes must decide whether to allocate funds equally to the development of public and private communications systems or to favor the development of one over the other.

Pool (1983) contends that there are four policy issues regarding the development of modern communications: (1) how much to invest; (2) the role to be played by public and private interests; (3) how much freedom to allow; and (4) what cultural level to promote. He points out

> . . . Communist and non-Communist countries make very different decisions on these matters. The practice in non-Communist countries is for development plans to provide large amounts for education of children and for the eradication of illiteracy. . . . The practice in Communist countries also is to spend heavily for literacy and

education but in addition to support expensive programs of exhortation addressed to adults (p. 243)

With regard to communications policy in Hungary, Szecsko and Fodor (1974, p. 22) note that since Hungary is "a transitional society whose people, after a political revolution, are living through revolutionary transformations of their culture and way of life, . . . mass communications must . . . play a larger part in the process of socialization—in making the new social values and norms accepted . . ."

Sandford (1976, pp. 203–9) comments on television in East Germany: "On both channels programming policy is informed by a determination to use the medium in the formation of a new socialist consciousness among the people." By law, two-thirds of all programming in the GDR must originate from non-Western sources. Sandford also notes that quite often the popular serials on East German TV feature a Western capitalist villain, such as a gangster or business tycoon. Among the programs broadcast in the GDR is one called the "Black Channel," which uses excerpts from Western television to illustrate the evils of capitalism.

There is little doubt that Eastern European television is used, at least in part, to shape the attitudes of the people towards socialism. But what of Western societies? This is somewhat more problematic in that there is substantial variation in the permissible uses of television among the Western industrial nations. The U.S., for example, is far more liberal than other nations in the amount of freedom allowed to advertisers, editorialists, and so on. The U.S. is also the only country in which television is almost wholly in private hands (though nominally regulated by a government agency).

Despite such variations in policy, it is clear that Western industrial nations use their TV systems, to some extent, as instruments to shape public opinion and promote conformity to some set of desired values and norms. In 1974 France instituted a new policy for television under which the three networks would be partially funded by the government and partially by commercial revenues. This "reform" followed a period of strict government control originating in the DeGaulle era. Still, Ardagh (1982, p. 407) points out that the traditional close relationship between government and commerce promotes the

---

[2] If variability in government styles, particularly within Marxist societies, were an important consideration, one would expect more variability in the outcome of the analysis. This is not the case. However, since within-group variation is not the topic of this research, my data are not designed to address it, leaving a potentially interesting topic for future research.

interests of business through television. He notes that consumer-advocate telecasts are monitored closely by the government and sometimes even suspended. Clearly, the government and business in France work together to promote consumerism.

Obviously, American television glorifies consumerism. Both programming and commercials advocate conspicuous consumption. Zwerdling (1985, p. 38) points out that mass advertising has contributed to the phenomenal growth of small local or regional companies, such as Beatrice Foods, to multinational status. In sum, both groups of nations have clear broadcasting objectives, whether it be the promotion of the "socialist man" or a better antiperspirant. Given this argument, the major difference between the two sets of nations, in terms of their application of television technology, lies on the underlying message: consumerism on the one hand, socialist ideals on the other. What remains to be explained is the difference in the growth and diffusion of the telephone.

The telephone in Western capitalist nations may also be an integral part of a consumer society. Lenski (1984, pp. 417–18) points out that collective action, as a means of attaining rewards previously only available to the wealthier classes, has been greatly enhanced by the "*unparalleled opportunities to communicate* [author's emphasis] among the less powerful and less privileged segments of society." This certainly includes the opportunity to use the telephone.

Oddly enough, the telephone was not originally envisioned as a means of casual interpersonal communication. Fischer (1983, p. 9) observes that the Bell company originally sought to sell the telephone as a business device. However, businesses were not ready customers, because the telegraph not only suited their needs, it established a written record of all transactions.

In the long run, the development of American telephony was in the interest of big business, and the Bell company went to great lengths to prove it. Yet, as Fischer (1983) points out, the phone was not initially seen as suitable or profitable for residential development. However, it was as a residential convenience that the telephone made its first major expansion. Clearly, its growth as a private means of communication was facilitated by the ever-growing wealth and power

of the nation and an ideology that permitted the free exchange of ideas.

The main characters in Solzhenitsyn's, *First Circle*, are political prisoners involved in a project to develop a device to identify speakers on the telephone through voiceprints, reflecting Marxist leaders' view of the telephone as a troubling technology. Boettinger (1977, p. 206) notes that Stalin quickly vetoed Trotsky's plan to develop a modern telephone system in Russia with the comment, "It will unmake our work. No greater instrument for counterrevolution and conspiracy can be imagined."

In sum, there is no basic difference in the application of television technology in Marxist and non-Marxist industrial nations. Both use it to encourage the development of certain attitudes and behaviors. The telephone, on the other hand, is problematic, at least in regimes wishing to maintain centralized control. While such countries could establish a system of monitoring telephone conversation, it would be prohibitively expensive. In the short term, it costs little to let the system stagnate or grow at a very slow pace. However, over the longer term, the ability to communicate quickly over long distances has important implications for societal development. The advanced Western industrial societies are shifting rapidly away from heavy industry into service-oriented enterprises. This movement is greatly enhanced by the free flow of information, through highly developed communications systems. Personal computers, for example, can access vast stores of information through telephone hookups with other computers. Clearly, such development will be impeded in those nations with less developed telecommunications.

Groth and Hunt (1985, pp. 123–36) postulate that Marxist governments, compared with non-Marxist nations, allocate relatively fewer resources to the development of interpersonal means of communication, such as telephones and mail service, than to public means. Using simple percentage comparisons and controlling for income, the authors found the general hypothesis was supported and that communications-system development was not highly correlated with income. The authors use multiple measures of public forms of communications. This inquiry, on the other hand, uses a ratio of telephones to televisions that allows direct comparisons between Marxist and non-Marxist nations. Moreover, Groth

Table 1. Rank Ordering of 31 Nations by Ratio of
         Telephones/Televisions (1975)

Nation	Ratio	Nation	Ratio
Switzerland	224.1	Italy	121.4
New Zealand	196.6	Spain	120.0
Sweden	188.2	United Kingdom	118.7
Portugal	185.5	Austria	114.9
Iceland	178.4	Belgium	111.9
Greece	176.2	W. Germany	103.6
Japan	174.8	France	97.4
Luxembourg	167.2	Ireland	74.0
Australia	165.1	Czechoslovakia	71.0
Denmark	147.4	Bulgaria	51.1
Canada	139.9	E. Germany	49.7
Netherlands	138.9	Romania	46.5
Norway	137.8	Hungary	44.6
Finland	127.4	Yugoslavia	42.4
United States	121.9	Poland	42.3
		USSR	30.6

Note: The ratio of telephones to televisions was
constructed as follows: Telephones/1,000 pop.
(1975)/televisions/1,000 pop. (1975) × 100. (Taylor and
Jodice [1983]). (V260/V248 × 100).

and Hunt were more interested in relative
resource allocation, while I am concerned
with ideology. Finally, my use of regression
analysis permits comparison of the strength of
the several explanatory variables.

FINDINGS

Table 1 shows the ratio of telephones to
televisions for the sample nations for 1975.
No overlap between the Marxist and non-
Marxist countries exists, and only the lowest
ratio of the non-Marxist nations (Ireland) and
the highest ratio of the Marxist nations
(Czechoslovakia) are comparable. Switzer-
land's ratio of 224.1 is more than three times
higher than that of Czechoslovakia (71.0), the
leading Marxist nation. The ratio in Ireland,
the lowest of the non-Marxist nations, is
nearly two-and-a-half times that of the Soviet
Union, the lowest of the Marxist group.

It could be argued that the disparity
between Marxist and non-Marxist societies is
a function of the lower standard of living in
Eastern Europe. To test this, I selected two
standard development measures: energy con-
sumption per capita and percentage of popu-
lation living in urban areas. I first matched
the Eastern European nations as closely as
possible with Western nations on each
measure (see Table 2). Then I placed the
variables in a regression equation excluding
other measures, such as GNP per capita,

because they are highly collinear with energy
consumption per capita.[3]

Another hallmark of advanced industrial
nations is high urbanization. Thus, it may be
argued that Western nations have a higher
ratio of telephones to televisions because they
are more urbanized. The evidence in Table 2
indicates that this argument is not tenable.
Three pairs of countries were matched exactly
on percent urban (Hungary-Sweden, Poland-
Norway, and Yugoslavia-Portugal). The dis-
parity in the ratio is greater than three to one
in all three. With regard to energy consump-
tion per capita, the disparities are much the
same.

Finally, diffusion of telephone and televi-
sion technologies could be a function of
population density. Since it is relatively more
expensive to provide telephone service to
sparsely populated areas, countries with low
population densities might be expected to
have fewer phones. However, while this
might explain low levels of each of the
individual technologies, it is less clear that it
should explain the *ratio* of telephones to
televisions. Given satellite technology, which
can be used in both television and telephone
transmission, the relative costs of providing
service should not differ appreciably. Also,
with regard to telephone systems, every
nation in the sample has trunk lines between
major urban areas. This being the case, it is
unlikely that the costs of extending telephone
service through traditional means (i.e., non-
satellite transmission) should be appreciably
higher than extending television service.

Nonetheless, in analyzing the data I have
included a measure of population density (see
Table 3). If population density is an important
factor in the diffusion of television and
telephone technologies, one should find a
high ratio in countries with high population
density and the opposite in low-density
countries. Table 3 demonstrates that this is

---

[3] Bollen (1979) indicates that the correlation can
be as high as 0.90. He further argues that using
energy consumption per capita in the regression
analysis avoids problems of exchange rates and
comparability often encountered with GNP/capita.
This collinearity is further documented by Frisbee
and Clarke (1979, p. 578), who, in the process of
constructing an index of development note that
". . . scholars who have relied on energy
consumption as a single-variable index seem to
have been on the right track."

Table 2. Ratio of Telephones/TVs for Marxist and Non-Marxist Nations, Controlling for Energy Consumption per Capita 1975, and Percent Urban 1975

Energy Consumption/capita:

Nation	Energy	Ratio	Nation	Energy	Ratio
	Cons/cap	Phone/TV		Cons/cap	Phone/TV
Czechos.	7,170	71.0	(No non-Marxist equivalent)		
E. Germany	6,484	49.7	(No non-Marxist equivalent)		
USSR	5,050	30.6	Norway	5,106	137.8
Poland	5,026	42.3	Denmark	5,018	147.4
Bulgaria	4,742	51.1	Finland	4,731	127.4
Rumania	3,812	46.5	Austria	3,724	114.9
Hungary	3,376	44.6	New Zealand	3,316	196.6
Yugoslavia	1,933	42.4	Greece	2,030	176.2
	$\bar{X}=4,699$	$\bar{X}=47.3$		$\bar{X}=3,987$	$\bar{X}=150.1$
	s.d. = 1,678	s.d. = 11.5		s.d. = 1,201	s.d. = 30.8

Percent Urbanization:

Nation	%Urban	Ratio	Nation	%Urban	Ratio
		Phone/TV			Phone/TV
USSR	36	30.6	W. Germany	35	103.6
GDR	24	49.7	Ireland	23	74.0
Poland	20	42.3	Norway	20	137.8
Hungary	28	44.6	Sweden	28	188.2
Czechos.	17	71.0	Finland	23	127.4
Yugoslavia	13	42.4	Portugal	13	185.5
Bulgaria	24	51.1	Denmrak	27	147.4
Rumania	25	46.5	Italy	29	121.4
	$\bar{X}=23.4$	$\bar{X}=47.3$		$\bar{X}=24.8$	$\bar{X}=135.7$
	s.d. = 7.0	s.d. = 11.5		s.d. = 6.6	s.d. = 38.7

Note: For ratio variable, see Table 1. Energy Consumption/capita 1975 (V120) Percent Urban (V280) (Taylor and Jodice [1983]).

not the case. With the exception of the Soviet Union, all of the Marxist countries have relatively high population densities and relatively low ratios of telephones to televisions. Conversely, Australia, Canada, and Iceland, with only five people per square mile, have relatively high ratios. Comparing Canada and the USSR, with similar population/land mass ratios, the Canadian ratio of telephones to televisions is more than four times greater.

Although population density does not predict the ratio of telephones to televisions, possibly a related variable, the average distance between people in a given area, might be a factor. Based on models suggested by Hawley (1950) and Mayhew and Levinger (1976), Nolan (1979) found that the average-distance variable had an effect on the size of government, as did the availability of modern communications and transportation technologies. While population density and average distance may appear to measure the same thing, population density changes exponentially with increased land area, while average distance changes logistically. Thus, if popula-

tion and land mass increase at the same rate, population density remains constant, while the average distance between individuals increases.[4] This is particularly important for countries with large land masses, such as the USSR, the US, Canada, and Australia. I included a measure of average distance in my regression analysis.

Finally, I included in the regression equation a dummy variable, "Marxist," with the Eastern European nations coded 1 and the remainder of the sample coded 0, to test the hypothesis that the nature of the regime accounts for the observed differences. The dependent variable is the ratio of telephones to televisions for 1975.

Tables 1, 2, and 3 demonstrate a disparity

---

[4] Average distance is a measure of the distance between all possible random pairs of individuals in a given area. Nolan (1979, p. 617) defines average distance as ".52 times the square root of the area of the system." Thus, it is a conceptual measure and cannot be expressed in terms of a meaningful metric.

Table 3. Population Density, Average Distance, and the Ratio of Telephones to Televisions (1975)

Nation	Pop. Density	Avg. Dist.	Ratio Phones/TVs
Netherlands	868	20.5	138.9
Belgium	835	15.5	111.9
Japan	775	186.0	174.8
W. Germany	647	124.5	103.6
United Kingdom	595	122.0	118.7
Italy	481	150.5	121.4
Switzerland	403	20.5	224.1
Luxembourg	359	1.5	167.2
Denmark	304	21.5	147.4
France	250	273.5	97.4[a]
Portugal	247	46.0	185.5
Austria	234	42.0	114.9
Spain	182	252.5	120.0
Greece	179	66.0	176.2
Ireland	114	35.0	74.0[a]
USA	60	4681.5	121.9[b]
Sweden	47	225.0	188.2[b]
Finland	36	168.5	127.4[b]
Norway	31	162.0	137.8[b]
New Zealand	29	134.5	196.6[b]
Australia	5	3843.5	165.1[b]
Iceland	5	51.5	178.4[b]
Canada	5	4988.0	139.9[b]
E. Germany	406	54.0	49.7[a]
Czechoslovakia	302	64.0	71.0[a]
Hungary	294	46.5	44.6[a]
Poland	283	156.5	42.3[a]
Romania	231	118.5	46.5[a]
Yugoslavia	216	128.0	42.4[a]
Bulgaria	205	55.5	51.1[a]
USSR	29	11201.0	30.6

Note: For ratio variable, see Table 1. Population Density (V104) (Taylor and Jodice [1983]).

[a] High density/low ratio of telephones/televisions.

[b] Low density/low ratio of telephones/televisions.

in the growth and diffusion of telephone and television technologies between the two sets of nations that cannot be satisfactorily explained by developmental factors. Univariate

analysis of the variables showed that energy consumption/capita, population density, and average distance were somewhat skewed. Thus, I used the natural log of these variables to produce a more normal distribution. I then put these logged variables into my regression equation.

Table 4 shows the means, standard deviations, and correlations for three groups: Marxist, non-Marxist, and the full sample. For the Marxist group, the population-density and average-distance measures are highly correlated with the ratio, while for the non-Marxist group they are not. I suspect that this is accounted for by the vast size of the USSR and by the small number of countries in the group. In other words, the USSR is clearly an outlier, in terms of size and population density, within the Marxist group. On the other hand, the effects of the large countries in the non-Marxist group are neutralized by the number of nations in the group (23).

The first model in Table 5 shows the result of a regression excluding the Marxist dummy variable. The significance of average distance and percent urban are increased, while that of population density remains about the same. However, the proportion of explained variance is reduced by half (.40) compared with the full model.

In the second regression model in Table 5, average distance, population density, and the Marxist/non-Marxist dummy variable are all significant, but the magnitude of the effect is greatest for the Marxist/non-Marxist variable. Marxist nation status decreases the logged ratio of telephones to televisions by a factor of 0.98. The range of the logged ratio runs

Table 4. Means, Standard Deviations, and Correlations for Marxist Group, Non-Marxist Group, and Full Model

Variable	Marxist		Non-Marxist		Full Model	
	Mean	(s.d.)	Mean	(s.d.)	Mean	(s.d.)
LRATIO75	3.83	(0.234)	4.94	(0.264)	4.66	(0.556)
LOGDENSITY	4.37	(0.812)	3.89	(1.637)	4.02	(1.471)
LOGAVDIST	1,478.00	(3,928.9)	679.65	(1,527.1)	885.68	(2,331.9)
PERCENT URBAN	23.38	(7.009)	38.09	(19.197)	34.29	(18.02)
MARXIST	1.00	(0.000)	0.00	(0.000)	N/A	
LNECC75	8.38	(0.414)	8.40	(0.573)	8.40	(0.529)

Correlation with Dependent Variable (significance in parentheses)

Variable		LRATIO75	LOG DENSITY	LOGAVDIST	PERCENT URBAN	MARXIST	LNECC75
	Marxist	—	.72 (0.04)	−.71 (0.04)	−.59 (0.12)	—	.37 (0.35)
LRATIO75	Non-Marxist	—	−.23 (0.29)	−.01 (0.96)	.06 (0.75)	—	−.08 (0.71)
	All	—	−.18 (0.34)	−.26 (0.17)	.32 (0.07)	−.89 (0.0001)	.01 (0.95)

Table 5. Regression of the Logged Ratio of Telephones/
Televisions (1975) on Marxist/Non-Marxist,
Logged Energy Consumption per Capita,
Percent Urban, Logged Population Density,
and Logged Average Distance (unstandardized
coefficients)

Dependent Variable: LNRATIO75			
	(Marxist Dummy Excluded)	(Full Model)	(USSR Excluded)
Regressors	Regression Coefficient	Regression Coefficient	Regression Coefficient
	(Standard errors in parentheses)		
Marxist		−0.98*** (0.12)	−0.96*** (0.12)
LNECC75	−0.070 (0.163)	−0.025 (0.087)	−0.022 (0.088)
%URBAN75	0.019** (0.006)	0.005 (0.004)	0.004 (0.004)
LNPOPDENS	−0.137* (0.072)	−0.0698* (0.0395)	−0.0685* (0.0395)
LNAVDIST	−0.229*** (0.065)	−0.093** (0.038)	−0.074* (0.043)
Constant	6.21	5.69	5.60
R^2	0.40	0.83	0.81
F (d.f.)	4.32 (4,26)	25.11 (5,25)	20.17 (5,24)

Note: For Ratio, Population Density, Percent Urban,
and Energy Consumption, see previous tables. Average
Distance (AVDIST) is computed from V103, (Taylor and
Jodice [1983]). All variables except Marxist (dummy
variable, where Marxist nations are coded 1, others
coded 0) and Percent Urban, were logged to reduce skew.
   * $p<.10$.
  ** $p<.05$.
 *** $p<.001$.

from 3.42 to 5.41, so a coefficient of this size
is highly significant. The logged value of
population density is not very large and is
significant only at the .10 level (−0.069).
Finally, a unit change in the logged value of
the average distance between individuals
lowers the logged ratio by a factor of 0.093,
which is not very important given the range of
values for the logged ratio (3.74–5.41).

Given the size of some of the sample
nations, the significance of population density
and average distance might be attributable to
outliers. Therefore, in the third model in
Table 4, I omitted the USSR, the largest
nation in the sample. The results are not much
different from the first model. The same
variables are significant, but average distance
is now only significant at the .10 level, rather
than at .05. I suspect that removing other
large nations from the sample would absorb
the significance of average distance alto-

gether. I conclude that the average distance
between people in societies is only a factor in
very large nations, where population is more
widely scattered. In both regressions, the fact
of being Marxist has the greatest effect on the
ratio of telephones to televisions.

CONCLUSION AND DISCUSSION

This research considered possible explana-
tions for the disparity in the ratio of
telephones to televisions between Marxist and
non-Marxist industrialized nations. On theo-
retical grounds, I argued that the type of
regime provides the best explanation. The
best alternative explanation involved techno-
economic development. Using descriptive
statistics and regression analysis, I found that
economic and demographic factors do not
satisfactorily explain the ratio disparities.

Two other alternative explanations for the
disparity were considered. First, the costs of
building and maintaining telephone and tele-
vision systems may differ. Unfortunately,
accurate and consistent data of this type are
not available. Also, the relative cost to the
consumer of telephones and televisions might
well be a factor, but, again, the data are not
available. The second is consumer prefer-
ence. In Western societies, both televisions
and telephones are readily available and
affordable for most people. Whether this is
the case in Marxist societies is not ascertain-
able, again because of the lack of data. Still,
given the magnitude of the Marxist effects
and the proportion of variance explained by
the variables included in the analysis (.83),
the addition of further variables would not
likely change the outcome significantly.
Clearly, there is a distinct difference between
Marxist and non-Marxist nations with regard
to the diffusion of telephones and televisions
that cannot be explained by economic or
demographic factors.

It would be interesting to analyze these data
differentiating between residential and busi-
ness/government telephones. I suspect that
there is approximate parity between Marxist
and non-Marxist nations in terms of the
diffusion of telephones in business and
government offices. The disparity in ratio of
telephones to televisions would be further
illuminated by comparing the diffusion of
residential telephones between the two sets of
nations. In fact, AT&T Longlines periodi-
cally publishes a volume entitled *Telephones*

Around the World, which contains such information; however, the amount of missing data makes it impossible to do a reasonable analysis.

The advent of computer technology has generated in modern nations a need for effective communications networks, which will play an ever more important role in international affairs. It will be of interest to see how Eastern Europe and the USSR, which have until now lagged behind in the development of telephone systems, will cope with the necessity of modernizing and expanding these systems to remain competitive with the West.

REFERENCES

Ardagh, John. 1982. *France in the 1980s: The Definitive Book.* New York: Viking Penguin.

Boettinger, Henry M. 1977. "Our Sixth and a Half Sense." Pp. 200–7 in *The Social Impact of the Telephone,* edited by Ithiel de Sola Pool. Cambridge, MA: MIT Press.

Bollen, Kenneth A. 1979. "Political Democracy and the Timing of Development." *American Sociological Review* 44:572–87.

Bower, Robert T. 1985. *The Changing Television Audience in America.* New York: Columbia University Press.

Fischer, Claude. 1983. "Educating the Public: Selling Americans the Telephone, 1876–1940." Unpublished Paper. Dept. of Sociology, University of California-Berkeley.

Frisbee, W. Parker and Clifford J. Clarke. 1979. "Technology in Evolutionary and Ecological Perspective: Theory and Measurement at the Societal Level." *Social Forces* 58:591–613.

Groth, Alexander J. and William R. Hunt. 1985. "Marxist-Leninist Communications Systems in

Comparative Perspective." *Coexistence* 22:123–136.

Hawley, Amos. 1950. *Human Ecology: A Theory of Community Structure.* New York: Ronald Press.

Lenski, Gerhard E. 1984. *Power and Privilege: A Theory of Social Stratification.* Chapel Hill, NC: University of North Carolina Press.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man.* New York: McGraw-Hill.

Mayhew, B.H. and R.L. Levinger. 1976. "Size and the Density of Interaction in Human Aggregates." *American Journal of Sociology* 82:86–110.

Nolan, Patrick D. 1979. "Area-Distance, Contact Technology and Administrative Intensity in Societies." *Social Forces* 58:614–29.

Pool, Ithiel de Sola. 1983. *Forecasting the Telephone.* Norwood, NJ: Ablex.

Sandford, John. 1976. *The Mass Media of the German Speaking Countries.* London: Oswald Wolff.

Solzhenitsyn, Alexandr. 1969. *The First Circle.* New York: Bantam Books.

Szecsko, Tamas and Gabor Fodor. 1974. *Communications Policies in Hungary.* Paris: UNESCO Press.

Taylor, Charles L. and David A. Jodice. 1983. *The World Handbook of Political and Social Indicators III: 1948–1977.* [MRDF]. Berlin: International Institute for Comparative Social Research. Ann Arbor: Inter-University Consortium for Political and Social Research.

U.S. Bureau of the Census. *Statistical Abstract of the United States,* edited by William Lerner. Washington, DC: U.S. Bureau of the Census. 1970, 1976, 1985 editions.

Zwerdling, Daniel. 1985. "The Food Monsters." Pp. 38–53 in *Crisis in American Institutions,* 6th ed., edited by Jerome Skolnick and Elliott Currie. Boston: Little, Brown.

URBAN BIAS, DEPENDENCE, AND ECONOMIC STAGNATION IN NONCORE NATIONS*

BRUCE LONDON
Florida Atlantic University

DAVID A. SMITH
University of California, Irvine

This paper contributes to the growing body of research that assesses the effects of patterns of urbanization on economic growth in less developed countries (LDCs). Regression analysis of data for 65–80 nations (including analyses of influential cases) suggests that both urban bias and dependency have independent, simultaneous, and negative effects on economic growth. Further analyses reveal a strong positive relationship between dependency and urban bias. These findings show that understanding urbanization and development in the Third World requires the consideration of both "external" (international) and "internal" (intranational) factors, and a combination of the insights gained from dependency/world-system, human-ecology, and modernization theories.

From Marx and Weber through Hawley and Castells, social theorists have postulated an ineluctable interpenetration of urbanization and other forms of social and economic development. In recent years, a dramatic shift of rapid urban growth to the less developed or Third World countries (LDCs), combined with their continuing poverty, has heightened interest in this issue. The combination of rapid urban growth with relative economic stagnation and poverty in LDCs challenges conventional assumptions about the relationship.

Early recognition of these trends led researchers to suggest that one of the major problems in LDCs is "overurbanization" (Davis and Golden 1954; Gibbs and Martin 1962). They argued that the populations of cities were increasing beyond the capability of their societies to cope with the human influx. But critics quickly pointed to theoretical and methodological problems in the overurbanization thesis (Sovani 1964; Kammerschen 1969; Hawley 1971; Castells 1977 [1972]). In subsequent years, the term *overurbanization* and the issues it addressed

became somewhat *passé* in academic discourse.

Renewed interest in overurbanization and urban/rural imbalance in Third World countries may be attributed partly to a seminal 1977 book by Michael Lipton. He argued that "urban bias" is a crucial cause of poverty in LDCs and that foreign involvement is not important. In fact, Lipton cited a small *positive* association between measures of "openness" to external trade and economic growth as evidence that international dependency cannot be the cause of Third World poverty (p. 81).

> As for the domestic-foreign polarity, it could seem odd that the UB (urban bias) argument attributes much rural poverty to rural-urban relations within LDCs, yet denies that much LDC poverty is attributable to LDC-DC relations. The reason for the latter is empirical: poverty and slow growth are statistically unrelated to involvement, or lack of it, in the world economy. (Lipton 1984, p. 157)

Instead, the real reason "why poor people stay poor" is that urban elites funnel an inordinate share of the resources of their societies into large cities, which have become "centers of power and privilege" (Gugler 1982, p. 188). The urban concentration of financial and human resources leads not only to rural-urban inequality but to a reduction of potential *aggregate* output of overurbanized LDCs.

Gugler (1982) used Lipton's notion of urban bias to reformulate the overurbanization thesis. He contended that urban bias and its accompanying rural exploitation are critical to

understanding underdevelopment in many Third World countries. In a cross-national analysis, Timberlake and Kentor (1983) supported the proposition that "the degree of overurbanization and changes therein inhibit economic growth" (p. 489).

Clearly, the renewed interest in overurbanization and urban bias has redirected attention to a crucial fact: aspects of Third World urbanization are often antithetical to national development. However, some analysts fear an overemphasis on overurbanization or urban bias as the principal explanations of stagnation. In particular, "the new urban sociology" (Walton 1981) and the emergence of the dependency/world-system paradigm have suggested that urbanization and development must be studied in the context of international political economy. Theories of the "dependent city" (Castells 1977 [1972]: chap. 3) and "peripheral urbanization" (Kentor 1981; Walton 1982) link LDC urban growth with incorporation into, and continuing subordination to, a capitalist world-economy dominated by Western core countries (see Chase-Dunn 1984 or Timberlake 1985, 1987). Viewed from this perspective, dependency, foreign investment penetration, and noncore world-system status (rather than urban/rural inequality or rapid urban growth) are the real sources of economic stagnation.

In a critique of the overurbanization thesis, Smith (1987) argued that *both* overurbanization and economic stagnation result from dependent status in the world-economy, and that, therefore, it is misleading to conceptualize stagnation as the outcome of only demographic inequalities. Timberlake and Kentor (1983) corroborated Smith's argument by demonstrating that dependence on foreign capital investment leads to overurbanization. A recent study by Bradshaw (1985) agreed.

This paper extends the research on the effects of rates and patterns of urbanization on economic growth in LDCs along several dimensions. Following the above arguments, our primary goal is to test the relative effects of urban bias and international dependence on economic stagnation in Third World nations. This requires an assessment of competing theories or explanations of LDC economic stagnation and thus responds to Bradshaw's (1987) and London's (1987) recent call for simultaneously testing hypotheses derived from divergent theories of urbanization and underdevelopment.

At a more general level, this work addresses the emerging debate over the relative importance of *intra*national versus *inter*national forces in determining patterns and rates of economic change in LDCs. Chirot (1981), for example, noted that dependency and world-system theorists shifted their analysis from the study of modernization towards international power relations: ". . . this neglects internal causes of social change just as surely as the earlier modernization theorists ignored the importance of world politics and unequal exchange between different economies" (p. 259). By rejecting the internal focus of conventional theorists, dependency/world-system proponents have tended to concentrate on a state's position in the world economy, downplaying undeniably important internal characteristics of noncore nations.

In this context, a number of recent studies have simultaneously examined both internal and external factors as independent variables in quantitative analyses (see especially Snyder and Kick 1979; London 1987; London and Williams, 1988). For our purposes here, we interpret urban bias as an intranational cause of economic stagnation, while dependency is an international determinant.

Finally, this study shifts the focus of empirical analysis from the concept of overurbanization to urban bias. Each concept stresses a different aspect of urbanization as the source of economic stagnation. Most indicators of overurbanization measure the degree to which urban *population growth* exceeds either general economic growth or the growth of nonagricultural labor-force opportunities (see Kentor 1981; Timberlake and Kentor 1983; Bradshaw 1985; London 1987). Urban bias, on the other hand, focuses on government fiscal (and other) *policies* that disproportionately benefit urban areas relative to rural areas, producing rural-urban income inequality and/or disparities in rural-urban standards of living (see Bradshaw 1987, p. 225). While some research has already been published that documents the relationship between overurbanization and economic stagnation. (Timberlake and Kentor 1983; Bradshaw 1985), this study uses two measures of rural-urban sectoral inequality (see below) to assess the impact of urban bias (rather than overurbanization) on LDC economic development.

DATA AND METHODS

The research is based on a regression analysis of cross-national statistical data drawn largely from Bornschier and Chase-Dunn[1] (1985, Appendix). With some variations (noted below), our analysis replicates and elaborates their model of economic growth. The elaboration introduces the sequential or hierarchical inclusion/exclusion of indicators of dependency and urban bias into their original equations. In other words, we conduct an analysis that, first, parallels a prominent study and then expands on that analysis to address the specific question at hand, i.e., the relative import of inter- versus intranational factors as predictors of economic stagnation.

The Dependent Variable

We use a more recent measure of the dependent variable used by Bornschier and Chase-Dunn (1985, pp. 63, 91, 103): the average annual growth rate of GNP per capita (percent), 1965–1984. Our source is the *World Development Report, 1986*, pp. 180–81. Bornschier and Chase-Dunn's measure of economic growth covered 1965–1977. However, given the frequent finding that the effect of dependency on economic and demographic processes is slow (cf. London 1987; Kentor 1981; Bornschier and Chase-Dunn 1985), we use the most recent measure

[1] Bornschier and Chase-Dunn include controls for the level of economic development in 1965 (GNP per capita, logged to correct for skewness, LYN1965) and the square of this term (YN²). These terms are included because economic growth is expected to be a function of initial level of development and because the relationship between initial level of development and subsequent growth has been found to be nonlinear. Including both level of development and its square as independent variables is the proper technique for specifying a curvilinear relationship. (For precedent, see Bornschier and Chase-Dunn 1985, pp. 76–77, 92; Weede 1980; Jagodzinski and Weede 1981; Nolan 1983). In our analyses, however, we did not find a curvilinear effect in any of the equations examined. Apparently, the relationship between initial level of development and subsequent economic growth is curvilinear only when core nations are included in the population under examination (e.g., as in the work of Bornschier and Chase-Dunn). Since we restrict our analysis to noncore nations, we will report equations that do not include the polynomial term.

available. Also, percentage growth rates are much less likely to be severely skewed than the absolute change in GNP/capita scores and, therefore, are less likely to be affected by heteroskedastic disturbances (Hanushek and Jackson 1977; Jackman 1980).

The Basic Model

We can only have confidence in our findings regarding dependency and urban bias if we first control for the effects of other potential determinants of economic growth. Bornschier and Chase-Dunn (1985) controlled for the following variables: foreign direct investment (FDI), gross domestic investment (GDI), level of economic development (LYN1965), level of exports (EXPORT), and size of the domestic market (SIZE). We include all of those variables in our analysis. For a detailed description and justification of these variables and the data, see Bornschier and Chase-Dunn (1985), especially pages 91–92 and 156–59.

Because the concepts of investment, savings, and level of development are integral to the neoclassical economic explanation of Third World economic growth, we ascertain whether dependency and urban bias have significant effects on economic growth net of controls for these factors (i.e., FDI, GDI, and LYN1965). This procedure makes our analysis quite congruent with Bradshaw's (1987) recent attempt to test arguments of modernization, urban bias, and dependency theories in a single cross-national study.

Independent Variables: Dependency

We measure "investment dependence," i.e., the penetration of an LDC's economy by multinational corporations ca. 1967 (LPEN). This measure is available for 103 nations in Bornschier and Chase-Dunn (1985, pp. 59–61) and indicates "transnational corporate penetration." It is computed according to the formula:

$$\text{PEN} = \sqrt{\frac{(K_{\text{FDI}})^2}{(K)(\text{POP})}}$$

where K_{FDI} represents the stock of capital controlled by foreign direct investment, K represents the total capital stock of the country, and *POP* represents the total population of the country. PEN is then logged (i.e.,

LPEN) to correct for skewness.[2] Although many different indicators of dependency are available, a measure of transnational corporate penetration gauges a crucial component of contemporary core-periphery relations, given the recent shift from "classical dependency," based on trade relations, to "the new dependency," based on investment and dependent industrialization (see Cardoso 1973; Cardoso and Faletto 1979; Evans 1979; dos Santos 1970; Bornschier and Chase-Dunn 1985, pp. 51–52. For a more detailed discussion of this point, see London 1987, pp. 33–34).

Independent Variables: Urban Bias

Ideally, a measure of urban bias *directly* taps the relative distribution of resources and investment to the cities versus the countryside. Unfortunately, good data on the spatial allocation of material inequality are not available. Therefore, like all other cross-national researchers, we use proxies. The literature suggests two potential measures. (The urban-bias indicators are the only independent variables not made available by Bornschier and Chase-Dunn.) The first is the measure of rural-urban disparity used by Bradshaw and based on arguments developed by Lipton (1977): the ratio of output per worker in nonagriculture to output per worker in agriculture (see Bradshaw 1987, pp. 230–31), measured for 1965 and logged to correct for skewness (LDISP65).

The second measure is a gini index of sectoral inequality in 1970 (GSECT70), which uses data on product per worker in the agricultural, service, and manufacturing sectors to measure "the imbalance between the rural subsistence farming sector and the urban commercial and manufacturing sectors of a dual economy characteristic of a developing country" (Taylor and Jodice 1983, p. 139). This is the same variable used by London (1987) to measure "rural adversity." However, urban bias and rural adversity are clearly cognate concepts. Since urban bias is accompanied by rural exploitation (Gugler 1982), it makes little difference whether one describes sectoral inequality as rural disadvantages relative to the urban sector (i.e., rural adversity) or urban advantages relative to the rural sector (i.e., urban bias). Note that sectoral-inequality data are available for 1960 and 1970, but not for 1965; however, most of our independent variables are measured as of 1965. We first examine the effect of 1970 sectoral inequality. Then, to make our temporal specifications more accurate, we estimate 1965 sectoral inequality (by averaging the 1960 and 1970 measures)[3] and redo the analyses with this estimate (GSECT65) used in place of the 1970 measure.

While both Bradshaw's and London's measures are based on data on product per worker, "disparity" is a ratio and "sectoral inequality" is a gini index. The gini index is the most commonly used measure of inequality (Allison 1978). It compares the actual distribution of productivity by sector with an "ideal" (i.e., perfectly equal) distribution by calculating the extent to which the actual distribution varies from the ideal. However, the ratio measure is explicitly advocated by Lipton (1977, 1984); we will examine both. The measures of disparity and sectoral inequality are highly, but not perfectly, correlated ($r = .84$, $N = 74$). This suggests that they tap somewhat different dimensions of the same process. Also, the urban bias measures are only weakly correlated with level of national overurbanization (see London 1987, p. 32): the percentage of a nation's population in cities of 100,000 or more divided by the percentage of a nation's labor force in the industrial sector, 1970 (with disparity, $r = .08$, $N = 63$, $p = .27$; with sectoral inequality, $r = .15$, $N = 74$, $p = .105$).

Analysis

Since our goals involve assessing both the *relative* importance of urban bias and dependency as predictors of economic growth and the *simultaneous* impact of the intra- and international factors on economic growth, a hierarchical regression strategy is appropriate. We first examine a regression equation that

---

[2] For a more complete discussion of this variable's derivation, see Bornschier and Chase-Dunn 1985, pp. 59–60.

[3] Using a constructed variable (the average of 1960 and 1970 sectoral inequality) might lead to measurement problems. It is possible that urban-rural inequality does not change over time in a gradual monotonic pattern. Therefore, the results based on this estimate should be interpreted with caution.

includes both the basic control variables suggested by Bornschier and Chase-Dunn and one of the measures of urban bias only. The next equation includes the basic variables with the measure of multinational corporate penetration, but excludes the measure of urban bias. A final equation includes the urban bias and dependency measures simultaneously. This hierarchical strategy is used for each measure of urban bias: disparity in equations 1–3 of Table 1; sectoral inequality, 1970 in equations 4–6; and sectoral inequality, 1965 in equations 7–9.

Countries Included

Sample size is always limited by the availability of data. Given the problems of sampling bias and generalizability and the requirements and limitations of regression analysis, one tries to maximize the number of cases available for analyses, despite the fact that any nation with missing data on even one variable is dropped from the analysis. Therefore, given our focus on Third World underdevelopment, the sample is defined as all noncore nations, or, alternatively, all nations classified as semiperipheral or peripheral in Bollen's (1983) revision of Synder and Kick's (1979) measure of world-system position.[4] Based on data availability on certain measures, the total number of cases included in the regressions varies from 65 to 80.[5] These sample sizes are large enough to

permit us to interpret regression results with some confidence.

FINDINGS

Table 1[6] presents the results of the regression of economic growth, 1965–1984 (GRORT84) on various combinations of control and independent variables. The findings are strong and unequivocal. Both urban bias (regardless of the measure used) and multinational corporate penetration have significant[7] nega-

the 1970 and 1965 measures of sectoral inequality, but not in the equations that include the measure of rural-urban disparity. Nations marked with a ° are included in the equations for 1970 sectoral inequality and disparity, but not in the equations for 1965 sectoral inequality. The nations are Algeria, Argentina, Benin, Bolivia, Brazil, Burma, Burundi[a], Cameroon, Central African Republic, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Finland, Ghana, Guatemala[b], Guinea[b], Haiti[b], Honduras, Hong Kong, India, Indonesia, Ireland,[b] Israel[b], Ivory Coast, Jamaica, Jordan[b], Kenya, South Korea, Liberia, Libya[c], Madagascar, Malawi, Malaya, Mali[b], Mauritania, Mexico, Morocco, Nepal[c], Nicaragua, Niger, Nigeria, Pakistan[c], Panama, Paraguay, Peru, Philippines, Portugal[b], Rhodesia (Zimbabwe), Rwanda, Saudi Arabia[c], Senegal, Sierra Leone[c], Singapore, South Africa, Spain, Sri Lanka, Sudan, Syria, Tanzania, Thailand, Togo, Trinidad, Tunisia[c], Turkey, Uganda, Upper Volta[b] (Burkina Faso), Uruguay, Venezuela, Zaire, Zambia.

[6] Quantitative cross-national studies rarely include correlation matrices, discussions of bivariate relationships, or consideration of the potential problem of multicollinearity. In analyses where Ns/samples change from equation to equation, it would be necessary to present a different matrix for each equation. The analyses reported in Table 1 are based primarily on three populations ($N = 65$; $N = 75$; $N = 68$). As it turns out, only one of the relevant independent variable intercorrelations in the sample of 65, three in the sample of 75, and two in the sample of 68, exceed $r = .50$ (and none reach the $r = .60$ level). Given these low to modest correlations, multicollinearity should not be a problem.

[7] In analyses of this type, where the units of analysis are large aggregates and where the number of cases is small or relatively small, the ratio of the unstandardized partial regression coefficient to its standard error is the most reliable guide to interpreting the "significance level" of coefficients (cf. Pedhazur 1982, pp. 242–43). Coefficients are considered significant if the unstandardized coeffi-

[4] Snyder and Kick (1979, pp. 1110, 1114) examined four types of international networks: trade flows, treaty memberships, military intervention, and diplomatic relations. They then "block modelled" these networks and found nine "blocks" of nations. These, in turn, were collapsed into the core, periphery, and semiperiphery categories of world-system theory. Bollen's (1983, pp. 473–76) revision is based on an analysis of partial regression plots of residuals. He discovered six outliers, argued that they represent a misclassification of countries by Snyder and Kick's procedures, and supported this argument with historical data. As a result, Spain, Portugal, and South Africa are reclassified from the core to the semiperiphery, and Taiwan, Iraq, and Saudi Arabia are moved from the semiperiphery to the periphery.

[5] The nations included in the analyses are listed below. Nations marked with an [a] are found in only those equations from Table 1 that include the 1970 measure of sectoral inequality. Those marked with a [b] are included in the equations that include both

Table 1. Regression of Average Annual Real Growth of Income Per Capita, 1965–1984 on Measures of Urban Bias/Sectoral Inequality, Multinational Corporate Penetration, and Selected Controls (reported coefficients are betas)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GNP per capital, 1965[a]	−.23*	.02	−.07	−.21	.09	−.02	−.23*	.18	.03
Foreign Direct Investment	.16	.29**	.22*	.11	.26**	.20*	.14	.32**	.25**
Gross Domestic Investment	−.03	−.01	.01	−.02	−.04	−.02	.10	.04	.08
Export	.29*	.27*	.37**	.23*	.22*	.32**	.11	.18	.21*
Size	.42**	.42**	.37**	.39**	.37**	.34**	.33**	.28*	.26*
Disparity, 1965[a]	−.34**		−.27**						
Sectoral Inequality, 1970				−.31**		−.24**			
Sectoral Inequality, 1965							−.36**		−.25**
Multinational Penetration, 1967[a]		−.43**	−.36**		−.43**	−.38**		−.52**	−.44**
R^2	.27	.29	.34	.22	.26	.30	.26	.32	.36
Adj. R^2	.20	.22	.26	.15	.19	.23	.18	.25	.28
N	65	65	65	75	75	75	68	68	68

[a] Logged to correct for skewness.
* B at least 1.5 times its standard error.
** B at least twice its standard error.

tive effects on economic growth, even when the effects of the intranational and international factors are considered simultaneously. The initially or separately observed negative effects of each on economic growth are not spurious; both processes exert independent direct effects on economic growth, net of the effects of the control variables (note the generally significant effects of FDI, SIZE, and EXPORT).[8]

Our findings lend considerable insight into the theoretical issues defined above. Specifically, they contradict both the extreme urban-bias perspective of Lipton, who claimed that "poverty and slow growth are statistically unrelated to involvement, or lack of it, in the world economy" (1977, p. 15), and Smith's

cient is at least 1.5 times the size of its standard error. This approximates a .10 level of significance and is frequently reported in quantitative cross-national research (cf. Weede 1980; Timberlake and Kentor 1983; Fiala 1983; Delacroix and Ragin 1981; Bornschier, Chase-Dunn, and Rubinson 1978; Jaffe 1985).

[8] Analyses were also done for the same dependent variable used by Bornschier and Chase-Dunn: economic growth, 1965–77. The ultimate results were essentially the same for this briefer span (i.e., both urban bias and dependency had simultaneous independent effects), but they were by no means as strong or unequivocal as those presented in Table 1. This difference suggests that dependency and urban bias do act slowly on the economies of LDCs. The specification of appropriate lag time is, therefore, quite important (cf. London 1987). (The 1965–77 results are available on request.)

(1987) suggestion that the negative effect of urban bias on economic growth may be spurious because both urban/rural inequality and material stagnation are the result of Third World dependence.

However, given the recent discussions and demonstrations of the importance of influential cases in analyses of this sort (Dietz, Frey, and Kalof 1987; Muller 1986), we must carry our analysis one step further. The discussion above suggests that equations 3, 6, and 9 in Table 1 present the most accurate assessment of the relationships under examination. We reexamine these equations for the sensitivity of the parameter estimates to one or more influential cases by following the procedure Muller (1986) used. He noted that "Cook's D is a summary measure of the extent to which a data point is influential" (p. 441). Using the regression diagnostics available in SPSS (see Hull and Nie 1981, pp. 94–121), we find a number of influential cases in each of these equations: Algeria, Saudi Arabia, and Singapore have high Cook's D scores in equations 3 and 6, while Algeria and Singapore are highly influential in equation 9. Further analysis shows these cases to be outliers (as shown by high studentized residual scores), with Algeria and Saudi Arabia the two most extreme outliers in equations 3 and 6, while Singapore and Algeria are the two most extreme outliers in equation 9. Moreover, Algeria and Saudi Arabia tend to have among the highest Mahalanobis' distance scores in equations 3 and 6, indicating unusual scores on the independent variables.

Table 2. Analysis of Influential Cases (from Equations 3, 6, and 9 of Table 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GNP per capita, 1965[a]	−.16	−.06	−.15	−.09	−.00	−.08	.08	.04	.08
Foreign Direct Investment	.36**	.22*	.36**	.32**	.20*	.32**	.35**	.24**	.35**
Gross Domestic Investment	.08	.01	.04	.04	−.00	.02	−.01	.10	−.03
Export	.11	.30*	.17	.09	.24*	.14	.12	.14	.18
Size	.34**	.37**	.34**	.32**	.34**	.31**	.30**	.26*	.29**
Disparity, 1965[a]	−.30**	−.23*	−.34**						
Sectoral Inequality, 1970				−.28**	−.20*	−.32**			
Sectoral Inequality, 1965							−.20*	−.23*	−.24*
Multinational Penetration, 1967[a]	−.38**	−.36**	−.37**	−.41**	−.38**	−.39**	−.51**	−.43**	−.50**
R^2	.43	.32	.43	.37	.28	.38	.39	.34	.40
Adj. R^2	.36	.24	.36	.30	.20	.31	.32	.27	.33
N	62	65	63	72	75	73	66	68	67

[a] Logged to correct for skewness.
  * B at least 1.5 times its standard error.
 ** B at least twice its standard error.

Table 2 presents the results of a regression analysis with these influential cases dropped from the respective equations (see equations 1, 4, and 7). In all but one case (GSECT65 in equation 7), both the urban bias and the penetration effects are enhanced by the case deletion. Also, R^2 increases in each case. One could argue that the consistently higher parameter estimates for penetration lend (modestly) greater support to dependency theory than to the urban-bias perspective. However, the most noteworthy implication is that case deletion does not change our substantive interpretations or conclusions in any way. Both the intranational and international factors are simultaneous, significant predictors of economic growth.[9]

We now consider "corrective action" for the influential cases: ". . .when outliers occur because one or a few cases have extreme values of the dependent variable, a relatively simple corrective procedure is to set a ceiling on the range of scores" (Muller 1986, p. 442). In our analysis, Singapore and Saudi Arabia have extreme values of the economic-growth measure. Singapore's growth rate of 7.8 percent, the highest in the population under consideration, is almost three standard

deviations above the mean, while Saudi Arabia's 5.9 percent is two standard deviations above the mean and third highest in the sample. The ceiling was set at 5.6 percent to reduce the scores of all cases that were two or more standard deviations above the mean. Equations 2, 5, and 8 present the results for the various full samples with this ceiling corrective in effect. The penetration effect remains strong and the urban-bias effect, while diminished somewhat, remains statistically significant. However Algeria (in all three equations) and Saudi Arabia (in equations 2 and 4 only) continue to be influential cases. (The influence of Singapore diminishes because, unlike the other cases, it had a high score only on the dependent variable, but no unusual scores on the independent variables.) This suggests that deleting these cases is appropriate. Equations 3, 6, and 9 present the results of the combined ceiling-corrective and case deletion. In these equations, R^2s rise to their highest levels, and the effects of both urban bias and multinational corporate penetration remain simultaneously strong. Overall, the analysis of influential cases in Table 2 mirrors the findings in Table 1. This increases confidence in the results of Table 1 (which have not been "excessively fitted"), and in the interpretations based on those results: both urban bias and dependency have strong, independent, direct, negative effects on non-core economic growth between 1965 and 1984.[10]

---

[9] We also examined the equations presented here (including the analysis of influential cases) with the two "city-states" of Singapore and Hong Kong deleted. Given the absence of rural populations in these nations, the effects of both measures of urban bias were reduced somewhat, although they remained statistically significant. The effect of penetration was essentially unchanged. Analyses excluding these cases yield the same ultimate conclusions as analyses that include them.

---

[10] While the analysis of influential cases did not alter our basic conclusions, the effect of outliers was much stronger in the analysis of 1965–1977

Table 3. Regression of Measures of Urban Bias (GSECT 70 and LDISP65) on Multinational Corporate Penetration and Selected Controls (reported coefficients are betas)

	(1) GSECT70	(2) GSECT70	(3) LDISP65	(4) LDISP65
GNP per capita, 1965[a]	−.46**	−.63**	−.31**	−.43**
Foreign Direct Investment	−.26**	−.27**	−.21*	−.23**
Gross Domestic Investment	.04	.13	.04	.03
Export	.39**	.21*	.33**	.28**
Size	−.10	−.06	−.20	−.14
Multinational Penetration, 1967[a]	.23*	.32**	.25*	.33**
R^2	.31	.36	.31	.33**
Adj. R^2	.26	.30	.24	.27
N	80	78	70	69

[a] Logged to correct for skewness.
  * B at least 1.5 times its standard error.
  ** B at least twice its standard error.

These observations point to the need to explore the relationship between urban bias and dependency. This is in keeping with Smith's (1987) suggestion that both overurbanization/urban bias and economic stagnation in noncore nations may well be results of dependent status in the world system. Some support for this argument has already been found in Timberlake and Kentor's (1983) discovery of a significant relationship between investment dependence and overurbanization and in London's (1987) documentation of a relationship between multinational penetration and rural adversity/sectoral inequality.

We now assess the relationship between multinational corporate penetration and both measures of urban bias in the context of controls for all of the "modernization" variables incorporated into the basic model of economic growth developed by Bornschier and Chase-Dunn (1985) and this study. The results in Table 3 reveal a significant positive relationship between penetration and urban bias, net of the controls for investment, savings, level of development, and so on.

In equation 1, sectoral inequality, 1970 is regressed on the measure of penetration and controls, while in equation 3 disparity is the dependent variable. Given the demonstrated usefulness of regression diagnostics, we analyzed influential cases. The deletion of

economic growth (see note 8). In that analysis (available on request), the simultaneous importance of urban bias and dependency was not clear until the effects of influential cases were specified. The potential importance of outliers and influential cases requires careful consideration in all quantitative cross-national research.

Saudi Arabia and Libya from equation 1 and of Libya from equation 3 proved appropriate. The results in equations 2 and 4 show the effect of multinational corporate penetration on both measures of urban bias is enhanced markedly, with the beta increasing from .23* to .32** in the GSECT70 equations, and from .25* to .33** in the LDISP65 equations.

DISCUSSION

This paper finds strong relationships among urban bias, dependence, and economic stagnation in noncore countries. With the appropriate lag, when considered both separately and in tandem, both urban bias and economic dependence are negatively related to economic growth. A complete analysis must, therefore, consider both influences simultaneously. This finding highlights recent suggestions that urbanization and development in the Third World cannot be fully understood unless analyses include both international and intranational factors and combine the insights of dependency/world-system analysis with those of human ecology and modernization theory (Nemeth and Smith 1985a, 1985b; Meyer 1986; Bradshaw 1987; London 1980, 1987). This seems to be the case largely because international political-economic factors shape the sort of intranational factors (e.g., levels of rural-urban disparity and sectoral inequality) that affect both patterns of urbanization and, in turn, economic growth in noncore nations (London 1987). The ultimate conclusion may well be that researchers need to self-consciously rid themselves of the sort of theoretical blinders that lead to the "categorical" analysis (Lenski 1966) of *either*

urban bias *or* dependency, of *either* intrana-
tional *or* international factors, of *either*
modernization *or* political economy. This is
the case precisely because these ostensibly
antithetical approaches are "related to each
other in a specifiable and theoretically
meaningful manner" (London 1987, p. 41).
Carefully *specifying* the theoretical articula-
tion between international political-economy
approaches and theories focusing on local and
national dynamics has become a pressing
prerequisite for further research progress.

REFERENCES

Allison, Paul D. 1978. "Measures of Inequality."
    American Sociological Review 43:865–80.
Bollen, Kenneth. 1983. "World System Position,
    Dependency, and Democracy." *American Socio-
    logical Review* 48:468–97.
Bornschier, Volker and Christopher Chase-Dunn.
    1985. *Transnational Corporations and Underde-
    velopment.* New York: Praeger.
Bornschier, Volker, Christopher Chase-Dunn, and
    Richard Rubinson. 1978. "Cross-National Evi-
    dence of the Effects of Foreign Investment and
    Aid on Economic Growth and Inequality: A
    Survey of Findings and a Reanalysis." *American
    Journal of Sociology* 84:651–83.
Bradshaw, York W. 1985. "Overurbanization and
    Underdevelopment in Sub-Saharan Africa: A
    Cross-National Study." *Studies in Comparative
    International Development* 20:74–101.
———. 1987. "Urbanization and Underdevelop-
    ment: A Global Study of Modernization, Urban
    Bias, and Economic Dependency." *American
    Sociological Review* 52:224–39.
Cardoso, Fernando H. 1973. "Associated-
    Dependent Development: Theoretical and Prac-
    tical Implications." Pp. 142–76 in *Authoritarian
    Brazil,* edited by A. Stepan. New Haven, CT:
    Yale University Press.
Cardoso, Fernando H. and Enzo Faletto. 1979.
    Dependency and Development in Latin America.
    Berkeley: University of California Press.
Castells, M. 1977 [1972]. *The Urban Question: A
    Marxist Approach.* Cambridge: MIT Press.
Chase-Dunn, Christopher. 1984. "Urbanization in
    the World-System: New Directions for
    Research." Pp. 111–20 in *Cities in Transforma-
    tion,* edited by M.P. Smith. Beverly Hills: Sage.
Chirot, Daniel. 1981. "Changing Fashions in the
    Study of the Social Causes of Economic and
    Political Change." Pp. 259–82 in *The State of
    Sociology: Problems and Prospects,* edited by
    J.F. Short, Jr. Beverly Hills, CA: Sage.
Davis, K. and H. Golden. 1954. "Urbanization
    and the Development of Pre-Industrial Areas."
    Economic Development and Cultural Change
    3:6–24.

Delacroix, Jacques and Charles Ragin. 1981.
    "Structural Blockage: A Cross-National Study
    of Economic Dependency, State Efficacy, and
    Underdevelopment." *American Journal of Soci-
    ology* 86:1311–47.
Dietz, Thomas, R. Scott Frey, and Linda Kalof.
    1987. "Estimation with Cross-National Data:
    Robust and Nonparametric Methods." *American
    Sociological Review* 52:380–90.
dos Santos, Teotonio. 1970. "The Structure of
    Dependence." *American Economic Review*
    50:231–36.
Evans, Peter. 1979. *Dependent Development: The
    Alliance of Multinational, State and Local
    Capital in Brazil.* Princeton, NJ: Princeton
    University Press.
Fiala, Robert. 1983. "Inequality and the Service
    Sector in Less Developed Countries: A Re-
    analysis and Respecification." *American Sociolog-
    ical Review* 48:421–28.
Gibbs, J. and W. Martin. 1962. "Overurbaniza-
    tion, Technology, and the Division of Labor:
    International Patterns." *American Sociological
    Review* 27:667–77.
Gugler, Josef. 1982. "Overurbanization Reconsid-
    ered." *Economic Development and Cultural
    Change* 31:173–89.
Hanushek, E. and J. Jackson. 1977. *Statistical
    Methods for Social Scientists.* New York:
    Academic Press.
Hawley, Amos H. 1971. *Urban Society: An
    Ecological Approach.* New York: Ronald Press.
Hull, C. Hadlai and Norman H. Nie. 1981. *SPSS
    UPDATE 7–9.* New York: McGraw-Hill.
Jackman, Robert W. 1980. "A Note on Measure-
    ment of Growth Rates in Cross-National Analy-
    sis." *American Journal of Sociology* 36:604–17.
Jaffe, David. 1985. "Export Dependence and
    Economic Growth: A Reformulation and Respe-
    cification." *Social Forces* 64:102–18.
Jagodzinski, Wolfgang and Erich Weede. 1981.
    "Testing Curvilinear Propositions by Polyno-
    mial Regression with Particular Reference to the
    Interpretation of Standardized Solutions." *Qual-
    ity and Quantity* 15:447–63.
Kammerschen, D. 1969. "Further Analysis of
    Overurbanization." *Economic Development and
    Cultural Change.* 17:235–53.
Kentor, Jeffrey. 1981. "Structural Determinants of
    Peripheral Urbanization: The Effects of Interna-
    tional Dependence." *American Sociological
    Review* 46:201–11.
Lenski, Gerhard. 1966. *Power and Privilege.* New
    York: McGraw-Hill.
Lipton, Michael. 1977. *Why Poor People Stay
    Poor: Urban Bias in World Development.*
    Cambridge, MA: Harvard University Press.
Lipton, Michael. 1984. "Urban Bias Revisited."
    Journal of Development Studies 20:139–66.
London, Bruce. 1980. *Metropolis and Nation in*

Thailand: The Political Economy of Uneven Development. Boulder, CO: Westview Press.

———. 1987. "The Structural Determinants of Third World Urban Change: An Ecological and Political Economic Analysis." *American Sociological Review.* 52:28–43.

London, Bruce and Bruce Williams. 1988. "Multinational Corporate Penetration, Protest, and Basic Needs Provision in Noncore Nations: A Cross-National Analysis." *Social Forces* 66:747–73.

Meyer, David R. 1986. "The World System of Cities: Relations Between International Financial Metropolises and South American Cities." *Social Forces* 64:553–81.

Muller, Edward. 1986. "Income Inequality and Political Violence: The Effect of Influential Cases." *American Sociological Review* 51:441–45.

Nemeth, Roger J. and David A. Smith. 1985a. "International Trade and World-System Structure: A Multiple Network Analysis." *Review* (Fernand Braudel Center) 8:517–60.

———. 1985b. "The Political Economy of Contrasting Urban Hierarchies in South Korea and the Phillipines." Pp. 183–206 in *Urbanization in the World-Economy,* edited by M. Timberlake. Orlando: Academic Press.

Nolan, Patrick. 1983. "Status in the World System, Income Inequality, and Economic Growth." *American Journal of Sociology* 89:410–19.

Pedhazur, Elazar J. 1982. *Multiple Regression in Behavioral Research: Explanation and Prediction.* New York: Holt, Rinehart, and Winston.

Smith, D. 1987. "Overurbanization Reconceptualized: A Political Economy of the World-System Approach." *Urban Affairs Quarterly.* December 23:270–94.

Snyder, David and Edward Kick. 1979. "Structural Position in the World System and Economic Growth, 1955–1970: A Multiple Network Analysis of Transnational Interaction." *American Journal of Sociology* 84:1096–1126.

Sovani, N.V. 1964. "The Analysis of Over-Urbanization." *Economic Development and Cultural Change* 12:113–22.

Taylor, Charles Lewis and David A. Jodice. 1983. *World Handbook of Political and Social Indicators,* 3rd ed. New Haven: Yale University Press.

Timberlake, Michael. 1985. *Urbanization in the World-Economy.* Orlando: Academic Press.

———. 1987. "World-System Theory and the Study of Comparative Urbanization." Pp. 37–65 in *The Capitalist City: Global Restructuring and Community Politics,* edited by M. Smith and J. Feagin. London: Basil Blackwell.

Timberlake, Michael and Jeffrey Kentor. 1983. "Economic Dependence, Overurbanization, and Economic Growth: A Study of Less Developed Countries." *Sociological Quarterly* 24:489–508.

Walton, J. 1981. "The New Urban Sociology." *International Social Science Journal* 33:374–90.

———. 1982. "The International Economy and Peripheral Urbanization." Pp. 119–36 in *Urban Policy Under Capitalism,* edited by N. and S. Fainstein. Beverly Hills: Sage.

Weede, Erich. 1980. "Beyond Misspecification in Sociological Analyses of Income Inequality." *American Sociological Review* 45:497–501.

THE USE OF SUBJECTIVE INFORMATION IN STATISTICAL MODELS*

RONALD ANGEL

Rutgers University

WILLIAM GRONFEIN

*Indiana University and
Purdue University at Indianapolis*

This paper examines problems that arise in using self-reports of subjective information as predictors in mathematical models. The analysis demonstrates that the social construction of such subjective information makes its use in comparative analysis problematic since both its "accuracy" and the outcome for which it is employed as a predictor are influenced by the respondent's culture and social location. We argue that subjective information is socially and culturally constructed and, by definition, endogenous in models in which self-reports are used as predictors. Methodologically, this requires that the endogenous status of subjective information be explicitly modeled when it is used as a predictor in structural models. As an illustration of the substantive consequences of group-specific response patterns, we examine the use of self-reported health status in comparing the health levels of Mexican-Americans, blacks, and non-Hispanic whites.

The emergence of survey methodology as the preeminent data-gathering technique of modern social science has brought with it a reliance on self-reports of subjective information. Recent investigations, however, have demonstrated that the "accuracy" and meaning of subjective reports are affected by cultural, social, and psychological factors (Angel and Thoits 1987; Fienberg, Loftus, and Tanur 1985a). In addition to the problems of distortion typical of self-reports of objective information, self-reports of subjective states such as well-being, self-concept, and general health present special problems of

interpretation, especially when such information is used to compare groups that differ in culture or social class.

This analysis examines the complex nature of subjective information and makes some suggestions concerning the use of subjective self-reports as predictors in statistical models. Our argument is that self-reports of internal subjective states cannot be assumed to be comparable across individuals, especially when such individuals differ in important social characteristics, although current practice tends to use subjective information in ways that explicitly or implicitly make just such an assumption. We also argue that such usage results in potential biases that can be discerned only if the social processes that generate subjective reports are explicitly modeled.

Let us begin by distinguishing between self-reports of what we term "subjective" and "objective" information. By *subjective* we mean those states that, like well-being, general health, attitudes, or pain, have no objective external referent. By *objective* we mean reports of income, arrests, hospitalizations, etc., that do refer to some objective external reality. It is well known that reports of objective statuses can be incorrect for any number of reasons including memory loss, telescoping, etc. (Fienberg, Loftus, and Tanur 1985a). On the other hand, reports of subjective states cannot be "wrong" in the

same sense because individuals are, by definition, the only judges of what constitutes happiness, satisfaction, health, etc., for themselves. Threats to the validity of such information arise primarily from the differential meanings that respondents read into questions concerning internal, often ambiguous, states and from differences in the interpretive and linguistic contexts within which they construe their answers. Reports of objective and subjective states, then, entail entirely different response tasks and are, for all practical purposes, essentially different entities. In this discussion we focus exclusively on the latter.

Recent work on the cognitive aspects of survey research reveals that subjective reports are complex products of the need dispositions, cognitive capacities, and interpretive procedures used by the respondent to generate reports concerning internal states (Angel and Thoits 1987; Fienberg, Loftus, and Tanur 1985a, 1985b, 1985c; Jabine, Straf, Tanur, and Tourangeau 1984). It is also clear that the evaluative criteria and cognitive apparatus that individuals use to interpret internal states and structure reports of those states are culturally and socially influenced (Angel and Thoits 1987). To the extent that social groups embody divergent cultural or community standards, similar states or conditions may be accorded quite different interpretations by individuals, and different indicators may be deemed relevant in the production of self-evaluations.

Reports of subjective states are, therefore, problematic when used in comparisons of individuals who differ significantly in those cultural and social factors that influence the meaning of subjective information. Yet, this difficulty is often ignored. Mathematical modelers commonly assume what is to be determined—that the social processes that generate subjective ratings in one social actor are the same as those that generate subjective ratings in another.

There are numerous examples of the use of subjective information in the social sciences. To illustrate our points concerning the use of subjective information, we have chosen a particularly important one, the use of self-ratings of health in the analysis of physician use. As do many other researchers, those who study the social correlates of health and illness depend heavily on respondents' subjective reports. Health is, after all, an essentially

latent and subjective state. The choice of physician use is merely illustrative and we mean our point to apply more broadly.

Studies of medical care use typically attempt to control for the "need" for medical care. In most multivariate analyses, need is controlled using some form of subjective health status (Andersen, Kravits and Anderson 1978; Berki and Ashencraft 1979; Ware, Davies-Avery, and Donald 1978; Wolinsky 1978). There are numerous versions of these health measures, but they all ask respondents to observe themselves, rate their health, and report this observation accurately. All self-reported health status measures are subject to biases of varying magnitude resulting from psychological and social factors that affect self-monitoring, awareness of symptoms or health limitations, and willingness to report socially undesirable conditions (Angel and Cleary 1984; Dohrenwend and Dohrenwend 1969; Mechanic 1979; Ross and Mirowsky 1984; Tessler and Mechanic 1978). Biases associated with variables such as gender, ethnicity, and income have also been commonly reported (Verbrugge 1985; Johnston and Ware 1976; Ware, Brook, Davies-Avery, Williams, Stewart, Rogers, Donald, and Johnson 1980).

There are clear empirical, as well as theoretical grounds, then, for questioning the uncritical use of self-reported health for purposes of group comparison. When self-reported health is used as a control for need in multivariate models, the impact of subjective health on medical care use is confounded with the effects of other factors, such as culture and social class, that simultaneously account both for perceived health and for use. To deal with these problems, or to at least detect their presence, researchers must explicitly consider the impact of the social factors that affect subjective health. Methodologically, this entails using a technique (e.g., multiple equation models or some similar procedure) in which the subjective report is treated as a dependent variable at one stage and subsequently employed as a predictor of an outcome at another.

We conduct analyses based on two data sets. The first analysis demonstrates that self-reports of health status are often incongruent with physicians' evaluations and that the degree of mismatch between the two varies with factors such as race, income, and education. The second analysis demonstrates

that multivariate models of physician use produce quite different results when self-reported health is treated as an exogenous variable that adequately reflects an individual's level of need and when it is treated as an endogenous variable produced by many of the same factors that determine use.

ANALYSIS ONE

In the first analysis, we employ the Health and Nutrition Examination Survey I (HANES I) carried out by the National Center for Health Statistics between 1971 and 1975 (United States Department of Commerce 1979). This data set is particularly useful for our purpose because it is one of the few large-scale studies that contains both subjective health ratings and a physical examination for each respondent. The data employed here refer to a sample of 6,913 respondents, aged 25–74, who received the most detailed physical examination. Our logic in this analysis is the following: if self-reported health is mainly a reflection of actual physical health status, demographic and socioeconomic variables should have a relatively insignificant effect on these reports, net of the physician's evaluation.

Subjective health is measured by a general health probe on which the respondent rated his or her health as excellent, very good, good, fair, or poor. The measure is coded negatively so that a higher score reflects poorer self-perceived health. As part of the examination, the physician coded any disease conditions present according to International Classification of Disease (ICD) guidelines (United States Department of Health, Education, and Welfare 1967). Our measure of objective health is based on this examination.

The physician rated the severity of each condition on a three-point scale as either minimal, moderate, or severe. We collapsed the individual ICD diagnoses into 12 broader categories[1] and assigned respondents a score in each category weighted by the physician's

---

[1] Based on the ICD codes, the following categories were created: infectious diseases, cancers and neoplasms, metabolic disorders, blood diseases, diseases of the nervous system, circulatory diseases, respiratory diseases, digestive diseases, gastrointestinal disorders, skin conditions, diseases of the musculoskeletal system, and symptoms not elsewhere classified.

evaluation of the severity of the most severe condition in that category. In addition, we calculated the number of separate diagnoses each respondent received. Respondents who had no diagnosis in a particular category received a score of 0 for that category. We then regressed respondents' subjective health score on the 12 condition categories and the total number of diagnoses to create a "standardized" or "expected" health score, which reflects the level of health an individual would report were her or his response consistent with the physician's evaluation.

Clearly, the results of a physical examination do not provide a completely unambiguous measure of health. Illness conditions may be overlooked or misdiagnosed, and physicians may apply different criteria in making their assessments. Nonetheless, it would be difficult to imagine a superior practical validity criterion. Even if there is error in the physicians' assessments, we have no reason to expect large systematic differences (relative to those affecting self-reported health) in patterns of diagnoses between types of respondents.

Several regressions of self-reported health on evaluated health and various sociodemographic and economic characteristics are presented in Table 1. The independent variables are introduced in a stepwise manner beginning with evaluated health status. This table reveals a significant association between subjective health and the physician's evaluation. Adding socioeconomic variables, however, results in appreciable increments to explained variance. Adding all eight variables (age, race, gender, education, employment status, low income, medium income, and marital status) results in a more than doubling of explained variance over that of the baseline model.

Race, age, education, employment status, marital status, and low income are all significantly associated with self-reported health net of physician evaluated health. These findings, then, strongly suggest that self-reported health is influenced by basic demographic and socioeconomic variables independently of its association with evaluated health status. Self-reported health, therefore, appears to be seriously confounded with other variables, such as education and race, that are commonly employed as controls in multivariate models of medical care consumption.

Table 1. Regressions of Subjective Health on Objective Health and Selected Demographic and Economic Variables (N = 6,093)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Evaluated Health	.384**	.279**	.246*	.246**	.233**	.231**
Female		.067*	.076*	.078*	−.032	−.027
Black		.496**	.313**	.314**	.323**	.309**
Age		.016**	.009**	.009**	.007**	.007**
Education			−.094**	−.094**	−.089**	−.079**
Not Married				−.010	.008	−.029**
Employment Status[a]					−.279**	−.255**
Middle income[b]						.040
Low income[c]						.277**
Constant	2.54**	1.69*	3.08*	3.08*	3.33**	3.00**
Adjusted R^2	.111	.161	.222	.222	.233	.240

[a] Major activity during past three months: Working = 1; Something else = 0.
[b] $7,000 to $14,999 = 1; otherwise = 0.
[c] $0 to $6,999 = 1; otherwise = 0.
* Significant at < .05.
** Significant at < .001.

ANALYSIS TWO

In the second analysis, we illustrate the potential consequences of different ways of using subjective health in prediction equations by modeling health in three different ways. We first use self-reported health "as is," and treat an individual's untransformed self-report as if it were an adequate proxy for need or objective health and enter it directly into a model predicting physician use. Then, based on the assumption that an individual's subjective self-report is a socially and culturally constructed representation of his or her social position, we employ instrumental variables to create two "expected" levels of health indices for each individual in the sample, and employ these indices in the use model.

The logic of the instrumental variable techniques is to employ the entire sample to determine the typical level of health that would be expected for an individual with a particular set of economic and demographic characteristics. While individual health levels differ greatly, in the aggregate, we would expect idiosyncratic variations to cancel out and differences not associated with socioeconomic factors such as income, sex, and education to represent group-response tendencies. This procedure is analogous to using standardized mortality rates to assess excess mortality (e.g., Kitagawa and Hauser 1973).

In this analysis, we focus on the impact of race and Mexican ethnicity on subjective health. Blacks, Mexican-Americans, and non-Hispanic whites differ in terms of social class and culture. In addition, Mexican-Americans differ in terms of degree of acculturation. If

culture and social class affect the meaning of subjective health, we would expect to be able to discern their impact by comparing these groups. Our analysis, however, is intended to be illustrative; our arguments do not refer exclusively to any specific cultural group.

This second analysis is based on a medical care use survey conducted in late 1975 and early 1976 by the Center for Health Administration Studies and the National Opinion Research Center of the University of Chicago (Aday, Andersen, and Fleming 1980). The universe for this sample is the total noninstitutionalized population of the United States. Individuals with episodes of illness, rural southern blacks, and persons of Spanish heritage living in the southwestern United States were oversampled. Detailed interviews were conducted with one adult and one child under 17 in each sample household. Here, Mexican-American refers to those who identified themselves as Chicano, Mexican, Mexicano, or Mexican-American. This analysis uses all adults between the ages of 18 and 64 and is based on a division of the sample into four subgroups: White, non-Hispanics (N = 2,941); Blacks (607); Mexican-Americans who took the interview in English (276); and Mexican-Americans who took the interview partially or totally in Spanish (201) (Aday, Chiu, and Andersen 1980).

In this survey, respondents were asked whether their health was, in general, excellent, good, fair, or poor; whether during the previous year their health had caused a great deal of worry, some worry, hardly any worry, or no worry at all; and whether they had

experienced pain very often, fairly often, occasionally, or not at all. Finally, respondents were shown a list of 22 common symptoms and asked if they had experienced any of them during the previous year. A subset of 20 nondental symptoms are used in this analysis. Forty physicians were asked to provide criteria for the severity of these symptoms. They indicated how many people out of 100 with a particular symptom, in various broad age categories, should consult a physician (Aday, Andersen, and Fleming 1980, p. 306). We use the median of the percentages provided by the physicians as a severity weight for each symptom.[2] To summarize the information contained in these four health measures, we standardized and combined them into a summary illness score (ILLNESS), which is negatively coded so that an increase in the score represents poorer health.[3] This four-item scale has an alpha of .82.

Table 2 presents five models estimated using the SAS SYSREG procedure (SAS 1979). The first two columns present the results of regressions predicting self-reported health status. The final three columns present the results of regressions predicting the natural logarithm of the number of home and office physician visits during the previous year. The number of visits is truncated at 20 to eliminate the influence of extreme outliers.

---

[2] The 20 symptoms employed here are cough, weakness, tired in the mornings, tired for weeks, headaches, rash, diarrhea, shortness of breath, aching joints, pain or swelling in the joints, backaches, weight loss, heart pain, indigestion, vomiting, sore throat, sneezing, bleeding, pains in the gut, infected eyes or ears. Test-retest Pearson correlations for the severity ratings by physicians indicated substantial consistency for the majority of symptoms, especially in the adult age range employed in this paper (see Aday, Andersen, and Fleming 1980, Appendix E, for a detailed discussion).

[3] This score has a mean of 0 and a standard deviation of 3.2. A principal-components factor analysis of these four items produced a single factor with an eigenvalue of 2.58. When standardized scoring coefficients were computed, each variable had a similar weight. Regressions using this weighted score produced results identical to those in the text except for the scaling of health. Because the metric of health is entirely relative, we have chosen to present results using the simplest possible scaling procedure, merely summing the standardized scores.

The variables listed along the left margin represent the predictors in each model. Blanks indicate that the particular predictor was not used in the model.

It is probable that there are errors in respondents' reports of physician's visits. We have no reason, however, to suspect that reports of visits are systematically biased as a function of race and ethnicity in the same way that we suspect reports of subjective health may be. In addition, we should note that we are addressing only the situation (which occurs quite frequently) in which subjective information is used to predict objective outcomes. Because the dependent variable is a report of an objective entity (the number of times one has been to the doctor), the fact that such reports are perfectible is crucial; there is no way of dealing with an inaccurate objective outcome except to improve its reporting. It would make little sense to adjust the ultimate dependent variable in an analysis in terms of the factors whose impact we are interested in assessing. As we shall demonstrate, however, there are other alternatives for subjective reports used as predictors.

Model 1 is the regression used to create our first expected health measure (*ILL). In this model, self-reported health status is regressed on the natural logarithm of family income, gender, education, age, marital status, being employed, being unable to work at all, the length of time since one's last physical exam, and the interval since one's last dental exam. Our choice of variables is directed by the following logic: being married, having higher education, and higher family income are associated with better health; increasing age and female gender are consistently found to be associated with reports of poorer health. Family size is included since the three racial and ethnic groups differ significantly in this characteristic, which may be related to health levels (Sweet 1978). Mexican-Americans have larger families, which may decrease health because of such factors as crowded living conditions. Being employed requires at least a minimal level of health, while the inability to work may indicate seriously compromised health status. Finally, the recency of one's last physical and dental examinations captures important aspects of health knowledge and behavior. This combination of variables comprises what we refer to as the instrumental variables for health status.

Race and ethnicity are not included in this

Table 2. Models Predicting Health Status and the Total Number of Home and Office Physician Visits During the Preceding Year ($N=3{,}915$, standard errors in parentheses)

	Health		Visits		
	Model 1 (*ILL)	Model 2 (**ILL)	Model 3	Model 4	Model 5
ILLNESS			.131** (.004)		
*ILL				.276** (.014)	
**ILL					.296* (.016)
Mexican-American (English interview)		−.394* (1.98)	−.008 (.053)	−.037 (.062)	.073 (.063)
Mexican-American (Spanish interview)		−1.726** (.241)	−.122 (.065)	−.283** (.075)	.216** (.083)
Black		.014 (.142)	−.052 (.038)	−.059 (.044)	−.063 (.045)
Female	.886** (.102)	.866** (.102)	.240** (.026)	.073* (.033)	.076* (.034)
Education	−.110** (.017)	−.148** (.018)	.022** (.005)	.035** (.006)	.050** (.006)
Age	.030** (.004)	.025** (.004)	−.003** (.001)	−.010** (.001)	−.009** (.001)
ln (family income)	−.236** (.072)	−.190* (.076)	.047** (.019)	.113** (.023)	.096** (.023)
Marital status	−.126 (.109)	−.019 (.110)			
Family size	.012 (.028)	.017 (.029)			
Physical exams	−.655** (.094)	−.547** (.094)			
Dental visit	−.068 (.097)	−.008 (.097)			
Employed fulltime	−.377** (.106)	−.268** (.107)			
Unable to work	4.651** (.246)	4.334** (.246)			
Private insurance		−.367** (.135)	.081* (.036)	.082* (.042)	.183* (.043)
Public insurance		.497** (.127)	.154** (.034)	.162** (.040)	.019** (.042)
South		.032 (.122)	−.103** (.030)	−.091* (.035)	−.105** (.036)
West		−.014 (.134)	−.016 (.036)	−.016 (.042)	−.013 (.043)
Central city		.169 (.112)	.068* (.030)	.077* (.035)	.032 (.036)
Rural		−.197 (.115)	−.057 (.031)	−.073* (.036)	−.018 (.037)
Doctors/1,000		.023 (.017)	.003 (.005)	.005 (.005)	−.001 (.005)
Regular source		.804** (.133)	.357** (.036)	.419** (.042)	.188** (.045)
Intercept	2.242** (.628)	1.792* (.697)	.131 (.178)	−.336 (.211)	−.250 (.215)
R^2	.214	.239	.297	.180	.170

* Significant at $<.05$.
** Significant at $<.01$.

model, so the expected level of health produced for each individual represents the level of health we would predict if we knew his or her age, education, sex, marital status, etc., but not race or ethnicity. The effect of this "expected" or "standardized" health measure, compared to that of the untransformed health status measure, will give us some idea of the impact of race and ethnicity on physician use through their effect on self-reported health.

Model 2 in Table 2 (**ILL) is the first stage of a two-stage least-squares regression (Hanushek and Jackson 1977). Two-stage least squares is used when an independent variable is correlated with the error term, the statistical result of the potential bias we suspect in our health measure. We use this standardized measure, in addition to our first (*ILL), because the statistical properties of two-stage least squares are well understood. In two-stage least squares, the suspect variable is first regressed on all the other variables in the equation in which it is to be used as a predictor (stage 2), plus those variables that are its unique predictors. The predicted values from this first stage, stripped of their association with the error term, are used as predictors in the second stage. In model 2 of Table 2, all the instruments for health status used in model 1 are entered into the equation. In addition, race, ethnicity, and the exogenous variables to be used in predicting physician use in stage 2 are entered to remove any confounding of their impact with that of health status in the second-stage equation. In this and subsequent models, non-Hispanic whites are the reference category and are omitted from the equations.

Once again, models 1 and 2 differ in that model 1 is based solely on health-related sociodemographic and economic characteristics and produces an expected health level independent of any effects attributable to race or ethnicity. Model 2 allows the influence of race and ethnicity to affect the expected health level. The coefficients for the instruments of health in models 1 and 2 do not differ significantly. Model 2, however, reveals that, although blacks are similar to non-Hispanic whites in their health reports, Mexican-Americans, especially those who took the interview in Spanish, report significantly better health levels (less poor health).

Although it might be argued that these findings reflect actual better physical health for Spanish-speaking Mexican-Americans than for other groups, there is sufficient evidence to indicate that Mexican-Americans suffer at least as much morbidity as non-Hispanic whites, making the possibility of better health unlikely (Angel 1984; Angel and Cleary 1984; Schreiber and Homiak 1981). An alternative explanation is that subjective estimates of health status by Spanish-speaking respondents in this sample are systematically biased. The following analysis examines the consequences of this potential bias in multivariate models of physician use.

Model 3 in Table 2 regresses the natural logarithm of physician visits on untransformed self-reported health (ILLNESS) and a number of predisposing and enabling variables commonly used in models of physician use (Aday, Fleming, and Andersen 1984). These include education, age, family income, private health insurance, public health insurance, having a regular source of care, region of the country, central city or rural residence, and the number of doctors per 1,000 inhabitants in the county of residence. None of the use models (3, 4, and 5) includes the unique predictors of health status (marital status, family size, having had a physical or dental examination within the past year, being employed fulltime, or being unable to work) because they were used to create the expected health scores and because we theorize that, net of their association with insurance and income, these variables affect physician use only through health.

In model 3, after controlling for self-reported health status and other demographic, economic, and access factors, the coefficients for all racial and ethnic differences are statistically insignificant. Using the untransformed self-report measure, therefore, leads to the conclusion that the differences in the average number of physician visits per year between the four racial and ethnic groups are due entirely to differentials in illness and the socioeconomic, demographic, and access factors. It also suggests that low rates of physician use by Spanish-speaking Mexican-Americans is a reflection of their generally superior health.

Model 4 is similar to model 3 except that self-reported health is replaced by our first predicted health score (*ILL). Recall that this standardized measure is based on what we have been calling the instruments of health, but not race or ethnicity. A number of

important differences emerge between models 3 and 4. In both models, non-Hispanic whites, blacks, and English-speaking Mexican-Americans are statistically indistinguishable from the reference group, non-Hispanic whites. This model indicates, however, that Spanish-speaking Mexican-Americans significantly *underutilize* physician services compared to other groups. Substantively, this model says that if Spanish-speaking Mexican-Americans had the same health level as others in the sample with similar health-risk profiles, they would use more physician services. The lower R^2s associated with models 4 and 5 (compared to model 3) result from the fact that the instrumental variable specification largely accounts for the simultaneity of illness and physician use.

Model 5 is similar to models 3 and 4, except that we employ the expected health measure that includes race and ethnicity and the other exogenous variables predicting physicians visits (**ILL). This model indicates that Spanish-speaking Mexican-Americans slightly overutilize physician services compared to non-Hispanic whites. This results from the fact that the ethnicity variable was allowed to influence the standardized score (**ILL) produced in stage one of the two-stage procedure (model 2). Model 5, then, must be interpreted in light of model 2, which reveals that the slight overuse of services by Spanish speakers in the second model (Model 5) is the result of a greatly suppressed level of reported illness. Comparing both equations (models 2 and 5), we have a better idea of how the self-report operates than is possible from a single equation model.

DISCUSSION

Both instrumental approaches lead to a rather different conclusion concerning the impact of ethnicity on physician use than does use of the untransformed subjective health. Using the untransformed score leads to the conclusion that all differences in use between groups can be attributed to differences in health and the other variables in the model. Results of the first instrumental variable specification (*ILL) directly illustrated the effect of biased reports by Spanish-speaking Mexican-Americans by removing the impact of group-reporting tendencies from the health-status measure. The two-stage least-squares procedure (**ILL) provided the same information,

this time by revealing significant differences in the impact of ethnicity on self-reported health in the first-stage regression. Either of the two instrumental techniques provides more information than does the direct use of self-reported health.

Of course, the choice of instrumental variables depends on theoretical considerations and the researcher's judgment. This is not a problem unique to two-stage techniques or to structural modeling, but one that applies to all statistical model building. At the very least, one can employ two-stage or other techniques in conjunction with OLS to examine the behavior of self-reported subjective information in multivariate specifications.

Clearly, self-reports of objective information can also be inaccurate because of such factors as recency, salience, and social desirability. Our main point, however, is that there is a fundamental difference between reports of objective entities, such as income, age, sex, or the number of times one has been to the doctor in the past year, and reports of subjective states for which there is no objective referent. Reports of objective information can be inaccurate for any number of reasons and it is technically conceivable to improve the accuracy of such reports. Reports of subjective information are, on the other hand, neither right nor wrong by definition, because the report is, for all intents and purposes, the reality.

We hold that the socially constructed nature of subjective information does not constitute distortion, but is rather an integral part of the reality itself. Although work in the psychometrics and cognitive structure of such information should have a high place on our research agenda, there are limits to the extent to which similar stimuli can be presented to respondents from very different social locations. The meaning of such stimuli are, in principle, different, depending on the interpretive context. We suspect, therefore, that there are inherent limits to the "accuracy" with which we can measure health status or other subjective states and that researchers must think seriously about the modeling of such information, in addition to its psychometric properties.

As we noted earlier, subjective information is used extensively in other areas of social science. For example, although there are behavioral indices of states such as depres-

sion, self-reports concerning symptoms of contentment or dysphoria are typically used to categorize persons as depressed. Although it is useful and proper to use self-reports of this type in some ways (e.g., to assess within-group associations between self-reported symptomatology and mental health-care use), their use to *rank* individuals from different groups in terms of degree of depression is, from our perspective, questionable.

Let us end by restating our argument that the appropriate use of subjective information in mathematical formalizations requires a greater appreciation of their socially constructed nature. The failure to do so introduces the very real risk of confusing those factors that generate the self-report with those that generate any outcome for which the self-report is used as a predictor. Our analysis illustrates this danger and proposes one way of dealing with the complex and often ambiguous nature of self-reports in multivariate models. Numerous limited-information and full-information techniques exist and are readily available to study complex structural relationships and to address subjective self-reports. Our analysis employed a rather simple two-stage procedure for illustration, but more complex models can be estimated.

Finally, in our opinion, the formal modeling of the socially constructed nature of self-reported subjective states is an inherently conservative strategy. If group-specific reporting tendencies do not affect subjective reports for a particular group, procedures such as those used here will not alter their impact. For blacks and English-speaking Mexican-Americans in this analysis, the two-stage procedure did not affect the impact of health on physician use. On the other hand, if social factors do significantly affect subjective reports, the two-stage procedure will reveal this, as in the case of Spanish-speaking Mexican-Americans.

There is little reason, then, for not treating subjective information endogenously. Such a treatment provides at least as much information as taking subjective self-reports at face value and provides a mechanism for discerning potential statistical and substantive problems with such variables. We might propose in conclusion that it is always necessary to begin with the assumption that significant differences in culture or in social status affect subjective information and to proceed as if

this were the case. Only empirical analyses can disprove this assumption.

REFERENCES

Aday, Lu Ann, Ronald Andersen, and G.V. Fleming. 1980. *Health Care in the U.S.: Equitable for Whom?* Beverly Hills, CA: Sage.

Aday, Lu Ann, G.Y. Chiu, and Ronald Andersen. 1980. "Methodological Issues in Health Care Surveys Of the Spanish Heritage Population." *American Journal of Public Health* 70:367–74.

Aday, Lu Ann, Gretchen Fleming, and Ronald Andersen. 1984. *Access to Medical Care.* Chicago: Center for Health Administration Studies.

Andersen, Ronald, Joanne Kravits, and Odin W. Anderson. 1978. *Equity in Health Services: Empirical Analyses in Social Policy.* Cambridge, MA: Ballinger.

Angel, Ronald. 1984. "The Costs of Disability for Hispanic Males." *Social Science Quarterly* 65:426–43.

Angel, Ronald and Paul D. Cleary. 1984. "The Effects of Social Structure and Culture on Reported Health." *Social Science Quarterly* 65:14–28.

Angel, Ronald and Peggy Thoits. 1987. "The Impact of Culture on the Cognitive Structure of Illness." *Culture, Medicine and Psychiatry* 11:23–52.

Berki, S.E. and Marie L. Ashencraft. 1979. "On the Analysis of Ambulatory Utilization." *Medical Care* 17:1163–81.

Dohrenwend, Bruce P. and Barbara S. Dohrenwend. 1969. *Social Status and Psychological Disorder: A Causal Inquiry.* New York: Wiley-Interscience.

Fienberg, Stephen, Elizabeth Loftus, and Judith Tanur. 1985a. "Cognitive Aspects of Health Survey Methodology: An Overview." *Milbank Memorial Fund Quarterly/Health and Society* 63:547–64.

———. 1985b. "Recalling Pain and Other Symptoms." *Milbank Memorial Fund Quarterly/Health and Society* 63:582–97.

———. 1985c. "Cognitive Aspects of Health Surveys for Public Information and Policy." *Milbank Memorial Fund Quarterly/Health and Society* 63:598–614.

Hanushek, Eric A. and John E. Jackson. 1977. *Statistical Methods for Social Scientists.* New York: Academic.

Jabine, Thomas, Miron Straf, Judith Tanur, and Roger Tourangeau, eds. 1984. *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines.* Washington, DC: National Academy Press.

Johnston, Shawn A. and John E. Ware, Jr. 1976. "Income Group Differences in Relationships

Among Survey Measures of Physical and Mental Health." *Health Services Research* 11:416–29.

Kitagawa, Evelyn and Philip Hauser. 1973. *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology.* Cambridge: Harvard University Press.

Mechanic, David. 1979. "Correlates of Physician Utilization: Why do Major Multivariate Studies of Physician Utilization Find Trivial Psychosocial and Organizational Effects?" *Journal of Health and Social Behavior* 20:387–96.

Ross, Catherine E. and John Mirowsky. 1984. "The Worst Place and the Best Face." *Social Forces* 62:529–36.

SAS. 1979. *SAS User's Guide, 1979 Edition.* Cary, North Carolina: SAS Institute, Inc.

Schreiber, J.M. and J.P. Homiak. 1981. "Mexican Americans." Pp. 264–336 in *Ethnicity and Medical Care,* edited by A. Harwood. Cambridge, MA: Harvard University Press.

Sweet, J. 1978. "Indicators of Family and Household Structure of Racial and Ethnic Minorities in the United States." Pp. 221–60 in *The Demography of Racial and Ethnic Groups,* edited by F.D. Bean and W.P. Frisbie. New York: Academic Press.

Tessler, Richard and David Mechanic. 1978. "Psychological Distress and Perceived Health Status." *Journal of Health and Social Behavior* 19:254–62.

United States Department of Commerce. 1979.

National Technical Information Service. *Health and Nutrition Examination Survey (HANES I) 1971–1975.* Hyattsville, MD.: National Center for Health Statistics.

United States Department of Health, Education, and Welfare. 1967. *Eighth Revision International Classification of Diseases, Adopted for Use in the United States.* Washington, DC: U.S. Government Printing Office.

Verbrugge, Lois. 1985. "Gender and Health: An Update on Hypotheses and Evidence." *Journal of Health and Social Behavior* 26:156–85.

Ware, John E., Jr., Allyson Davies-Avery, and Cathy A. Donald. 1978. *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. V. General Health Perceptions.* Santa Monica, CA.: The Rand Corporation. R-1987/5-HEW.

Ware, J.E., Jr., R.H. Brook, A. Davies-Avery, K.N. Williams, A.L. Stewart, W.H. Rogers, C.A. Donald, and S.A. Johnston. 1980. *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. 1, Model of Health and Methodology.* Santa Monica, CA: The Rand Corporation, R-1987/1-HEW.

Wolinsky, F.D. 1978. "Assessing the Effects of Predisposing, Enabling, and Illness-Morbidity Characteristics on Health Service Utilization." *Journal of Health and Social Behavior* 19:384–96.

ASA Journals

Current and Forthcoming Articles

Teaching Sociology
Vol. 16, No. 2 (April 1988)

Articles on Teaching about Women and
Gender Roles

Moving Our Minds: Studying Women of Color
and Reconstructing Sociology, *Margaret L.
Anderson*

Mainstreaming Gender: Some Thoughts for the
Nonspecialist, *Laura Kramer* and *George T.
Martin, Jr.*

Transcending a Patriarchal Past: Teaching the
History of Women in Sociology, *Mary Jo
Deegan*

Other Articles

Teaching Sociology in the Humanist Tradition;
Brad Lowell Stone

"Computerization" of a Sociology Department:
The Experience at Ball State, *Jerry G. Bode
and William S. Johnson*

Sociology and Writing Across the Curriculum:
An Adaptation of the Sociological Journal,
Barbara C. Karcher

Notes

By its Right Name: The Relevance of Poetry for
Sociology, *James D. Miley*

What Can We Do About Essay Exams? *Patricia
Bizzell and Royce Singleton, Jr.*

Transforming the Sociology of the Family: New
Directions for Teaching and Texts, *Maxine
Baca Zinn and D. Stanley Eitzen*

Readings from Zaniello's Across-the-Disciplines:
Text as Illustrations of Basic Sociological
Concepts, *Joan Ferrante*

Illustrating the Classical Experiment, *Richard A.
Zeller*

Teaching Sociological Methods with Limited
Computer Resources, *G. David Curry*

Book Reviews

H. L. Tischler, P. Whitten, and D. E. K. Hunter,
Introduction to Sociology. Mostafa M. Noury

J. W. Coleman and D. R. Cressey, *Social
Problems,* 3rd ed. John F. Biter

T. Williams and W. Kornblum, *Growing up
Poor.* Louis J. Gesualdi

A. J. Cherlin and F. F. Furstenberg, *The New
American Grandparent: A Place in the Family,
A Life Apart.* Ann Oakes

M. B. Zinn and D. S. Eitzen, *Diversity in
American Families.* David L. Elliott

J. H. Turner, *The Structure of Sociological
Theory* 4th ed. Charles T. Griffin

R. J. Parelius and A. P. Parelius, *The Sociology
of Education,* 2nd ed. Paul Lindsay

S. K. Kachigan, *Statistical Analysis: An
Interdisciplinary Introduction to Univariate and
Multivariate Methods.* John B. Williamson

D. Nachmias and C. Nachmias, *Research
Methods in the Social Sciences,* 3rd ed. John
Seem

F. E. Hagan, *Introduction to Criminology:
Theories, Methods, and Criminal Behavior.*
Louis Howard Narcez

C. H. McCaghy and S. A. Cernkovich, *Crime in
American Society,* 2nd ed. Jack N. Rose

A. C. Twaddle and R. M. Hessler, *A Sociology
of Health,* 2nd ed. Norman R. Layne, Jr.

D. L. Miller, *Introduction to Collective Behavior.*
R. George Kirkpatrick.

G. S. Jowett and V. O'Donnell, *Propaganda and
Persuasion.* Philo C. Wasburn

Z. Luria, S. Friedman, and M. D. Rose. *Human
Sexuality.* Alfred A. Clarke, Jr.

A. Chapman, ed., *Feminist Resources for
Schools and Colleges: A Guide to Curricular
Materials.* Brenda D. Phillips

L. Furr, C. Howery, C. Bordieri, and E. Kain,
*Teaching Marriage and Family: Print and
Visual Resources.* Mary Jo Neitz

C. S. Green, *Guidelines and Resources for
Assessing Your Sociology Program.* H. Paul
Chalfant, Jeanne Ballantine

J. E. Stice, ed., *Developing Critical Thinking
and Problem-Solving Abilities.* Louis E.
Anderson

C. Meyers, *Teaching Students to Think Critically.*
P. J. Baker

D. Wallace, *Introduction to Information Society:
Text/Workbook with "Hands On"
Microcomputer Exercises,* 3rd ed. Ramona L.
Ford

Film and Video Reviews

Killing Us Softly and Calling the Shots. Jeffrey
Chin

*Heroes and Strangers: A Film About Our
Fathers.* Thomas J. Sullivan

Finnish American Lives. Tim Curry

The AIDS Movie. Edward L. Kain

How to Say No to A Rapist and Survive. Henry
W. Fischer, III

The Onion Field. A. Javier Trevino

Sociological Theory
Vol. 6, No. 1 (Spring 1988)

Principles of Theoretical Analysis, *Guillermina
Jasso*

DEATH NOTICES

◆

Received in 1987

ASA receives death notices of its members on an unsystematic basis. Sometimes obituaries are sent and published in *Footnotes*. In order to provide a more permanent basis of this information, death notices will be published annually in the June issue of *ASR*. Where possible, the date of death is given and the issue of *Footnotes* where the obituary appeared. ASA will appreciate additions and corrections to this list.

Cressey, Donald R., July 21, 1987, *FN*, December, 1987

Dorse, Alvin C., February 7, 1987, *FN*, Aug., 1987

Flynn, Charles, January 11, 1987, *FN*, May, 1987

Haferkamp, Hans, no date

Hess, Dick, on December 25, 1986

Fliegel, Frederick, September 11, 1987, *FN*, December, 1987

Kimbrough, Emory, Jr., January 8, 1987, *FN*, October, 1987

Lantz, Herman R., May 27, 1987, *FN*, October, 1987

LaPiere, Richard T., on February 2, 1988, *FN*, March, 1987

New, Peter, December 30, 1985

Peters, Anne K., February 25, 1987, *FN*, August, 1987

Queen, Stuart, September 28, 1987, *FN*, December, 1987

New & selected titles in sociology

TALES OF THE FIELD

On Writing Ethnography

JOHN VAN MAANEN

Van Maanen offers a survey of the narrative conventions associated with writing about culture—the matter-of-fact realistic report, the confessional tale of the participant-observer, and the dramatic vignette of the impressionist style—and analyzes the strengths and weaknesses of the various styles.

Paper $7.95   190 pages
Library cloth edition $25.00
Chicago Guides to Writing, Editing, and Publishing

THE TRANSFORMATION OF OLD AGE SECURITY

*Class and Politics
in the American Welfare State*

JILL QUADAGNO

Grounded in historical research and informed by social science theory, this study reveals how public assistance grew from colonial-era poor laws, locally financed and administered, into a massive federal bureaucracy.

Cloth $27.50   272 pages   6 halftones, 1 map

GOD'S CHOICE

*The Total World of a
Fundamentalist Christian School*

ALAN PESHKIN

"[God's Choice] provides an invaluable record of contemporary fundamentalist culture at a time when most reporting on the subject is animated by hostility."—Katie Leishman, *Atlantic*

Paper $12.95   349 pages

BECOMING AN EX

The Process of Role Exit

HELEN ROSE FUCHS EBAUGH

With a Foreword by Robert K. Merton

Ebaugh's pathbreaking inquiry into the phenomenon of becoming an ex reveals the profundity of this basic aspect of establishing an identity in contemporary life. Drawing on interviews with 185 people, Ebaugh explores a wide range of role changes, including ex-convicts, ex-alcoholics, divorcees, ex-nuns and ex-cops.

Paper $10.95   272 pages   3 line drawings
Library cloth edition $27.50

PROFESSION OF MEDICINE

*A Study in the
Sociology of Applied Knowledge*

ELIOT FREIDSON

"A landmark in the sociological analysis of the professions in modern society."—Ron Miller, *Sociological Quarterly*

Paper $14.95   440 pages
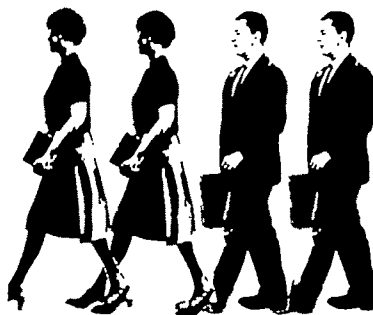
PROFESSIONAL POWERS

*A Study of the
Institutionalization of Formal Knowledge*

ELIOT FREIDSON

"This is a stimulating and well-written book which opens up a new perspective on the professions as well as contributing to existing debates."—David Podmore, *Times Higher Education Supplement*

Paper $12.95   241 pages

THE UNIVERSITY OF CHICAGO PRESS 5801 South Ellis Avenue Chicago, IL 60637

VI

SOCIOLOGICAL METHODOLOGY 1988

CLIFFORD CLOGG, Editor

- *Panels and Cohorts: Design and Model in the Study of Voting Turnout,* Otis Dudley Duncan and Magnus Stenbeck.
- *The Life History Calendar: A Technique for Collecting Retrospective Data,* Deborah Freedman, Arland Thornton, Donald Camburn, Duane Alwin and Linda Young-DeMarco.
- *Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence,* Mark Granovetter and Roland Soong.
- *A Method for Analyzing Backward Recurrence Time Data on Residential Mobility,* Nazil Baydar and Michael White.
- *Specification and Estimation of Continuous State Space Hazard Rate Models.* Trond Petersen.
- *The Analysis of Cross-Classified Categorical Data from Complex Sample Surveys,* J.N.K. Rao and D. Roland Thomas.
- *Some Models for the Multiway Contingency Table with a One-to-One Correspondence Among Categories,* Michael E. Sobel.
- *A Stabilized Newton-Raphson Algorithm for Log-Linear Models for Frequency Tables Derived by Indirect Observation,* Shelby J. Haberman.
- *An Application of Item Response Theory to the Measurement of Depression,* Nora Cate Shaeffer.
- *Evaluating a Multiple-Imputation Method for Recalibrating 1970 U.S. Census Detailed Industry Codes to the 1980 Standard,* Donald J. Treiman, William T. Bielby, and Man-tsun Cheng.
- *Causality in the Social Sciences,* Margaret M. Marini and Burton Singer.
- *Exploring Causal Structure with the TETRAD Program,* Clark Glymour, Ricard Scheines, and Peter Spirtes.
- *Causal Inference and Path Analysis,* Paul W. Holland.
- Discussion by Edward Leamer.

ASA Members: $30
Non-members: $35
Institutions: $40

American Sociological Association
1722 N Street, NW, Washington, DC 20036;
(202) 833-3410

Don't Miss This Valuable Resource!

Cumulative Index
of Sociology Journals

The Cumulative Index of Sociology Journals, published in early 1987, covers the 1971-1985 period. Compiled by Judith Lantz of Carbondale, it contains both an author and keyword or subject index. Books reviewed in any of the journals that are part of the Index are included in both the author and subject listings. The following ASA-sponsored publications are included in the 780-page Cumulative Index: American Sociological Review, The American Sociologist, Contemporary Sociology, Journal of Health and Social Behavior, Social Psychology Quarterly, Sociological Methodology, Sociological Theory, and Sociology of Education.

In addition, both the American Journal of Sociology and Social Forces have been incorporated into the index. Their inclusion represents the Association's first, and very preliminary, attempt to put together a single index of all major sociology journals. Future editions will include a larger number of journals.

The Cumulative Index is available to ASA members for $37.50, to non-members for $48.50 and to institutions for $65.00. Send prepaid orders to: ASA Executive Office, 1722 N Street, N.W., Washington, D.C. 20036. Act now to get this valuable new resource!

Please send me_____ copy (copies) of The Cumulative Index of Sociology Journals.

I have enclosed my check for_____, made out to ASA.

Name _____

Membership # _____

Address _____

_____

City _____ State _____ Zip _____

Please return this form and your check to the ASA Executive Office, 1722 N Street NW, Washington, DC 20036.